Journal of Experimental Education

Volume V | September, 1936 | Number 1

THE VERBAL ACCOMPANIMENT OF THE EDUCATIONAL FILM—THE RECORDED VOICE vs. THE VOICE OF THE CLASSROOM TEACHER[1]

John Elmore Hansen
University of Wisconsin

The sound motion picture has received considerable attention recently as an educational aid, and educators sense its great possibilities. It is important, however, that its possibilities and limitations be treated analytically and that enthusiasm for that which is novel be not allowed unduly to influence educational practice. Sound motion pictures fall into two classes: (1) the *bona fide* sound picture, in which natural sound is recorded and reproduced as an essential element of that which is portrayed; (2) the strictly talking picture, which has been photographed as a silent picture and to which has been added merely an explanatory lecture.

There should be no question concerning the educational value of the former. The use of the latter involves certain pedagogical questions, namely: (1) Should a formal verbal explanation accompany the presentation of pictorial materials or should these concrete particulars be translated into verbal forms by the pupils with the aid of the teacher? (2) What form should the verbal accompaniment take, if a verbal accompaniment is desirable? Should a complete explanation of everything that transpires in the picture be given, or should suggestive questions merely be raised and the answers be worked out by the pupils? (3) If a running commentary with the film presentation, regardless of its nature, is desirable, can it be presented more effectively through the medium of the recorded voice and the sound projector, that is, presenting the picture as a regular talking picture, or can the identical verbal continuity be presented as effectively by the teacher simultaneously with the projection of the picture through a silent projector?

The purpose of the study reported herein is to compare the effectiveness of the verbal continuity accompanying the talking picture when presented mechanically through the medium of the recorded voice and the sound motion projector, and when presented by the teacher simultaneously with the picture projected through a silent projector. Since most of the sound motion pictures available for educational purposes at the present time are strictly talking pictures with no natural sound, the answer to this question should have considerable value to the educational public.

Previous Studies

During the past few years several motion picture studies have been made which have touched either directly or indirectly upon this question.

In 1932 C. C. Clark[2] of New York University experimented with sound motion pictures, silent motion pictures, and classroom lecture-demonstrations to determine their relative effectiveness as a means for conveying concrete information and for stimulating and maintaining interests. He concluded that sound films of the lecture type in which all explanations are given by an unseen speaker are not as effective as identical silent films in which all explanations are given by captions.

In the spring of 1932 W. F. Einbecker[3] performed a series of experiments in the Deerfield Shields Township High School, Highland Park, Illinois, in which the following forms of

[1] The writer wishes to express his appreciation of the kind cooperation and guidance extended by Dr. A. S. Barr, Professor of Education at the University of Wisconsin, in the conduct of this experiment. Thanks are also due to Principal R. B. Woodworth, of the Roosevelt Junior High School, Fond du Lac; to Superintendent Roland A. Klaus of the Edgerton Public Schools; and to the teachers in these schools who so kindly participated in the study.

[2] C. C. Clark, "Sound Motion Pictures as an Aid in Classroom Teaching," *School Review*, XL (November, 1932), 669–81.

[3] W. F. Einbecker, "Comparison of Verbal Accompaniments to Films," *School Review*, XLI (March, 1933), 185–92.

verbal accompaniment were used: (1) presentation of a silent film without captions or oral comments of any kind; (2) presentation of a silent film with captions; (3) presentation of a silent film with captions and with oral comments by the teacher; (4) presentation of a silent film without captions but with oral comments by the teacher; (5) presentation of a talking picture. Einbecker concluded in part that it is immaterial whether the classroom teacher or the speaker in the talking picture presents the verbal continuity.

In 1934 Leon H. Westfall[4] reported a study in which a series of silent films and a series of sound films were each presented with several forms of verbal accompaniment. He concluded in part: (1) that for films of the science or expository type, originally constructed for use with sound accompaniment, a lecture that is furnished with the film and read by the teacher is significantly superior to the same material printed on the film or to an explanation which the teacher prepared from materials furnished with the film, but (2) that a mechanically produced lecture is superior to any other form of verbal accompaniment.

With the rather limited and contradictory information which now exists regarding the relative effectiveness of the several methods of presenting verbal information with the educational motion picture it would seem that further evidence is desirable.

Subjects

The subjects employed in this investigation consisted of approximately 415 pupils, including the 350 pupils in the seventh grade of the Roosevelt Junior High School of the Fond du Lac Public Schools and the 65 pupils in the tenth grade of the Edgerton Public Schools.

The pupils at Fond du Lac were divided into two groups of 170 matched pairs, matched first according to I.Q. and M.A. Immediately following the pre-test for each film they were more closely matched according to I.Q., M.A., and scores on the pre-test, which resulted in a considerable reduction in the numbers of pairs actually used in each study. The same procedure was followed at Edgerton with thirty original matched pairs. In effect this amounted to four separate studies in each city, since each picture presentation was treated as a separate study with the groups re-matched before each picture presentation.

The chief facts concerning the subjects are given in Tables I and II.

Table I

Motion Pictures, Number of Subjects, Mean I.Q., Mean M.A., and Mean Scores on T_1 for the Teacher Groups and Sound Groups Employed at Fond du Lac

Motion Pictures	No. of Subjects	I. Q. Averages	M. A. Averages	T_1 Averages
The Frog				
Teacher Group	102	102.83	158.04	61.08
Sound Group	102	103.22	157.71	60.98
Spiders				
Teacher Group	117	103.22	156.74	53.95
Sound Group	117	102.91	156.11	53.77
Fungus Plants				
Teacher Group	117	103.62	157.44	39.80
Sound Group	117	103.23	156.77	40.17
Plant Growth				
Teacher Group	113	103.28	156.99	54.57
Sound Group	113	102.43	156.18	54.84

[4] Leon H. Westfall, *A Study of Verbal Accompaniments to Educational Motion Pictures.* New York: Columbia University, 1934.

TABLE II

MOTION PICTURES, NUMBER OF SUBJECTS, MEAN I.Q., MEAN M.A., AND MEAN SCORES ON T_1 FOR THE TEACHER GROUPS AND SOUND GROUPS EMPLOYED AT EDGERTON

Motion Pictures	No. of Subjects	I. Q. Averages	M. A. Averages	T_1 Averages
The Frog				
Teacher Group	26	98.9	167.9	78.8
Sound Group	26	97.46	168.8	79.11
Spiders				
Teacher Group	26	100.07	169.4	67.3
Sound Group	26	98.5	170.1	67.3
Fungus Plants				
Teacher Group	23	97	169.1	67.26
Sound Group	23	97.5	165.7	66.87
Plant Growth				
Teacher Group	24	98.7	171.5	72.2
Sound Group	24	99.1	168.6	72.9

The rotated group method was used, each group having two films presented by the mechanical sound method and two by the silent-film-plus-teacher method. This served still further to equalize the groups receiving instruction by the two methods.

Five teachers participated as readers with the silent films.

MATERIALS OF INSTRUCTION

The materials of instruction consisted of the following four Erpi talking pictures: Plant Growth, Fungus Plants, The Frog, and Spiders; these are all talking pictures, consisting of a silent film accompanied by a verbal explanation presented mechanically by an off-stage voice.

To insure that the verbal continuity presented by the teacher would be identical in each case with the verbal continuity accompanying the talking picture, the spoken continuity of the film was recorded on a dictaphone and transcribed. The teacher in each case was presented with the typed continuity twenty-four hours before he was to present the film to his group. Three of the teachers reported reading the typed explanation with the film once, and two that they had gone through it twice with the film before actually presenting it to their respective groups.

TESTS

The four tests, one for each picture, consisted of twenty separate multiple choice questions of five items each, all, or none, or any of which might be correct. So that each test might truly be said to consist of one hundred items. Each test was given as a pre-test twenty-four hours before the film was presented in each case and again as a final test immediately following the film presentation. The reliability coefficients of the four tests when given as pre-tests were: .86, .937, .963, .824, as determined by split-test correlation and corrected by the Spearman–Brown formula. The reliability coefficients of the same tests when given as final tests were: .89, .937, .89, .90. No attempt was made to determine the validity of the tests, but care was exercised to make them cover every item of importance presented in the verbal continuity accompanying each film, this verbal continuity for each film being the same for both methods of presentation.

PRESENTATION OF SUBJECT MATTER

At Fond du Lac each of the four pictures was presented to the control group and to the experimental group in the school auditorium because each of the two groups was too large to be accommodated in a classroom. The same projector was used for both groups, with the exception that the sound was shut off for the teacher presentation of the verbal explanation.

When each group was assembled a two-minute explanation of the purpose of the study was given. This was followed by a

three-minute introduction to the picture, this introduction being identical for both types of presentation. Each picture ran approximately eleven minutes.

For the teacher presentation a different teacher was used for each picture at Fond du Lac. There were four teachers of elementary science in the seventh grade in this school, and these teachers participated in the study.

At Edgerton, since the tenth grade biology classes were taught by one teacher, the same teacher presented the verbal continuity with all four pictures. Since the groups here consisted of approximately thirty pupils each, the regular classroom was used for the picture presentation for both groups. Otherwise the presentation followed the same procedure as at Fond du Lac.

Results

The mean scores for the various Fond du Lac groups are presented in Table III and for the Edgerton groups in Table IV.

At Fond du Lac the mean gains for the teacher presentation for the pictures, The Frog, Spiders, Fungus Plants, and Plant Growth, were 17.19, 23.18, 23.77, and 9.55, respectively; and for the recorded voice 15.87, 21.85, 25.97, and 7.0, respectively. This represents a small advantage in gains for the teacher group over the recorded sound group for the first, second, and fourth pictures, while

Table III

Mean Scores and Standard Deviations for the Teacher and Sound Groups on Tests I and II; the Gains Between Tests; the Differences in Gains Between Groups; and the Standard Errors of the Differences at Fond du Lac

	T_1 Averages	T_2 Averages	T_{2-1} Averages	Difference in Gains	S. D. of Diff.
The Frog					
Teacher Group	61.08	78.27	17.19	1.32	1.116
σ(dis.)	10.536	12.044	11.64		
σ(av.)	1.043	1.192	1.15		
Sound Group	60.98	76.85	15.87		
σ(dis.)	10.11	9.77	8.78		
σ(av.)	1.043	.967	.869		
Spiders					
Teacher Group	53.95	77.13	23.18	1.33	1.41
σ(dis.)	14.99	13.855	12.45		
σ(av.)	1.38	1.258	1.15		
Sound Group	53.77	75.62	21.85		
σ(dis.)	14.136	13.317	14.07		
σ(av.)	1.303	1.23	1.30		
Fungus Plants					
Teacher Group	39.80	63.57	23,77		
σ(dis.)	14.92	12.788	17.51		
σ(av.)	1.373	1.177	1.619		
Sound Group	40.17	66.14	25.97	2.2	1.59
σ(dis.)	17.35	12.06	19.388		
σ(av.)	1.604	1.11	1.79		
Plant Growth					
Teacher Group	54.57	64.12	9.55	2.55	1.19
σ(dis.)	9.333	10.157	9.13		
σ(av.)	.874	.955	.859		
Sound Group	54.84	61.84	7.00		
σ(dis.)	9.38	10.583	10.67		
σ(av.)	.878	.992	1.00		

TABLE IV

MEAN SCORES AND STANDARD DEVIATIONS FOR THE TEACHER AND SOUND GROUPS ON TESTS I AND II; THE GAINS BETWEEN TESTS; THE DIFFERENCES IN GAINS BETWEEN GROUPS; AND THE STANDARD ERRORS OF THE DIFFERENCES AT EDGERTON

	T_1 Averages	T_2 Averages	T_{2-1} Averages	Difference in Gains	S. D. of Diff.
The Frog					
Teacher Group	78.8	90.27	11.47	0.62	1.465
σ(dis.)	6.49	5.37	6.053		
σ(av.)	1.27	1.05	1.189		
Sound Group	79.11	89.96	10.85		
σ(dis.)	5.92	3.94	4.93		
σ(av.)	1.16	.77	.967		
Spiders					
Teacher Group	67.3	90.4	23.1	6.0	1.62
σ(dis.)	6.04	5.54	6.43		
σ(av.)	1.185	1.09	1.26		
Sound Group	67.3	84.4	17.1		
σ(dis.)	5.99	7.04	6.01		
σ(av.)	1.175	1.38	1.179		
Fungus Plants					
Teacher Group	67.26	78.43	11.17		
σ(dis.)	6.27	6.68	8.62		
σ(av.)	1.307	1.39	1.79		
Sound Group	66.87	81.07	14.2	3.03	1.587
σ(dis.)	7.35	6.59	6.72		
σ(av.)	1.53	1.37	1.40		
Plant Growth					
Teacher Group	72.2	79.33	7.13	1.59	1.86
σ(dis.)	5.28	8.49	5.538		
σ(av.)	1.078	1.73	1.13		
Sound Group	72.9	78.46	5.54		
σ(dis.)	6.82	6.6	5.568		
σ(av.)	1.39	1.34	1.136		

for the third picture the advantage is in favor of the recorded voice or sound group. However, a comparison of these differences in gains with the standard error of the difference[5] in each case indicates that the differences are not statistically significant.

At Edgerton the mean gains for the teacher presentation for the same pictures were 11.47, 23.1, 11.17, and 7.13, respectively; and for the recorded voice 10.85, 17.1, 14.2, and 5.54, respectively. Here again the differences in gains are too small to be statistically significant, except in the case of the picture, Spiders, where the net average gain for the teacher presentation is 6.0 points higher than that for the recorded voice, which is 3.7 times the standard error of the difference.

On the whole the difference in the results from the two methods of presentation cannot be considered significant.

[5] Standard Error of the difference in gains between groups of matched pairs has been computed as equal to:

$$\sqrt{\sigma M_x + \sigma M_y - 2r_{xy}\,\sigma M_x\,\sigma M_y}$$

σM_x = S. D. of Av. of x gains
σM_y = S. D. of Av. of y gains
r_{xy} = correlation between individual x and y gains
x = individual gains for group with teacher presentation
y = individual gains for group with sound recording presentation
M_x = mean of x gains
M_y = mean of y gains

Remarks

Although no vote was taken, it was apparent by their attitude that the pupils favored overwhelmingly the sound or recorded voice method of presentation. They expressed their delight in being assigned to the group that was to see and hear the talking picture and had difficulty in repressing their disappointment when assigned to the group to which the teacher was to read the verbal continuity. This enthusiasm for the recorded voice method of presentation was probably due almost entirely to the fact that the sound motion picture was a novelty whereas the silent educational motion pictures were used quite regularly in their classroom instruction. The teachers who participated expressed their disfavor with the method of presenting the verbal explanation as employed in the study and emphasized their desire to present the silent film by their own method, the method that they had adopted as a result of several years of experience in the use of educational films.

At Fond du Lac, where the presentation was made in the auditorium, a number of pupils complained that they could not hear the teacher satisfactorily. The voice from the sound projector, however, could be heard distinctly by all of the pupils. At Edgerton, where the presentation was made in a classroom, the pupils seemed to have no difficulty in hearing either method of presentation distinctly.

Conclusions

The results of this study seem to indicate that the verbal explanation accompanying an educational picture of the talking picture type can be presented as effectively by the classroom teacher as by the medium of the recorded voice and the sound motion picture projector. It is important though to bear in mind that this conclusion does not necessarily hold for all teachers or for all motion pictures. The personality and the voice of the teacher will greatly determine the effectiveness of the teacher presentation. On the other hand, a poorly recorded voice and poor room acoustics may impair the effectiveness of the recorded voice. It is probably fair to conclude that for ordinary classroom teaching the teacher's voice is as effective as the recorded voice, while in the large auditorium the greater volume and the tone control afforded by the sound projector make the recorded voice more effective. Although the teacher's presentation of the verbal explanation may not be as closely synchronized with the pictorial presentation as that of the recorded voice, it is doubtful whether this has much if any effect, especially if the teacher is familiar with the subject matter being presented.

It should be apparent, of course, that much saving in film and projector cost could be afforded if the present talking pictures were made available in the form of silent pictures accompanied by the verbal explanation in printed form. Also, the silent picture lends itself more readily to use by the individual teachers in their classrooms, and the greater possibility of its use at the opportune moment will probably make for more effective teaching.

THE INTEGRATED VERSUS THE NON-INTEGRATED USE OF MOVING PICTURES IN THE CLASSROOM

C. D. JAYNE
Central State Teachers College, Wisconsin

I. THE PROBLEM AND ITS SIGNIFICANCE

The purpose of this investigation is to gain some information in regard to the relative teaching value of two commonly used, fundamentally different, methods of using educational films in the classroom. By one method a film is used incidentally, sometime before or after the teaching of the unit to which it pertains. In this article this procedure will be referred to as the non-integrated method, or Procedure I. By the second method a film is used at the time the unit of work to which it pertains is being taught, in such a manner that the film becomes an integral part of the teaching unit. This method will be referred as the integrated method, or Procedure II.

Authorities in the field of visual education seem to agree that the integrated method of using films is the ideal method. They are not agreed, however, as to which is the more practical method, since it costs from 40% to 50% more to book films so that they can be integrated with the regular classwork. The basis for this difference of opinion is a disagreement as to the relative educational value to be derived from the integrated and non-integrated use of films. Some visual education authorities and users of visual aids take the position that films unless integrated with the regular classwork are of very questionable teaching value and that money spent on non-integrated films is largely wasted from an educational viewpoint. Others argue that films are of great educational value whether integrated with the regular classwork or not and that the difference in the products of learning from the two methods is not great enough to justify the increased cost of the integrated method. In actual school practice it seems that the field is rather evenly divided between the two procedures.

When there is such a wide divergence in theory and in practice on such an important point it seems obvious that it should be taken out of the realm of opinion and submitted to careful experimentation. The object of this investigation is to make a beginning in that direction.

More specifically, the purpose of this investigation is to gather data from experimentation which will throw some light upon the answer to the following questions:

1. How much more information do children gain from an educational film when it is studied intensively, as an integral part of a learning unit, than when it is shown incidentally sometime before the formal class study of the unit of work to which it pertains?

2. Do children gain more information from a learning unit (as distinguished from the film alone) when a moving picture is used as an integral part of the unit, than when the moving picture is shown sometime before the start of the unit?

3. Is the difference in the amount of information gained by either method, as compared with the other, large enough to be a significant factor in determining the method which should be used in presenting films?

Turning now to a more detailed account of the specific aspects of the problem covered by this study, it should be noted that the study consists of what are essentially two separate investigations, each in a different subject matter field, with different types of films, and with pupils of different grade placement. Thus the specific purpose of one investigation is to compare the mean gain in information of sixth grade geography classes from the study of the units "Austria" and "Hungary", when a film over each unit is shown incidentally two weeks before the start of the unit, with the mean gain made by the same classes over the same units when the film was used as an integral part of the teaching unit. The specific purpose of the second investigation is to compare the mean gain in informa-

tion made by ninth grade general science classes from a study of the units "Hot Air Heating" and "Simple Machines", when a film over each unit is shown incidentally two weeks before the start of the unit, with the mean gains made by the same classes over the same units when the films are used as an integral part of the teaching unit.

II. Experimental Procedure

The pupils participating in the study consisted of 79 sixth grade pupils and 80 ninth grade pupils in the public schools of Stoughton, Wisconsin. In the sixth grade the two classes used had been carefully equated on the basis of I.Q. at the beginning of the school year and so were known to be of practically equal capacity as measured by the Terman Intelligence Test.[1] The ninth grade classes were not equated on the basis of I.Q., but since the rotated group form of experiment was used any difference in I.Q. was "rotated out" and thus did not affect the validity of the results of the study.[2]

The units of work employed in the sixth grade geography classes consisted of one on Austria and one on Hungary. The book material for these units was from *Nations Beyond the Seas* by Atwood and Thomas[3] and was supplemented by some reference material. The teaching films "Austria" and "Hungary" were used with their respective units.

The units of work used in the ninth grade general science classes consisted of one on Hot Air Heating and one on Simple Machines. The book material for these units was from *Every Day Problems in Science* by Pieper and Beauchamp.[4] The teaching films "Hot Air Heating" and "Simple Machines" were used with their respective units.

The tests used as instruments of measurement in this study were four in number,[5] one for each unit of work. They were all of the objective type. Each test consisted of two parts, the first covering the film material, the second the book material not in the film. The gains from seeing the film alone were measured by the first section of the test; the total gains from the study of the unit were measured by the entire test. In constructing the tests an effort was made to cover the book material as adequately as the film material.

The method used in making the test should insure curricular validity. The film portion of each test was made to cover the facts presented by the picture as listed after viewing the film at least three times. The film portion of the geography tests consists of 40 items, and that portion of the general science tests consists of 34 items. It should be noted that many of the items in the film test were covered by the text as well as by the film.[6] The portion of the tests covering the book material alone was based upon an analysis of the facts presented in the textbooks used in the Stoughton Schools. The unit test in geography consisted of 60 items. The unit test in general science consisted of 52 items.

While recognizing that understandings, attitudes, and appreciations should be fundamental objectives in the teaching of units such as these, this study is limited to a consideration of only the informational changes produced in the pupils. Consequently all of the test items are factual in character and attempt to measure nothing but the level of informational achievement.

The rotation method of experiment was used as is indicated in Table I. This method seemed most desirable for several reasons. First, whatever differences there may have been in the general ability of the geography classes or of the general science classes would be eliminated as a factor influencing the results of the experiment because each film procedure would be affected in like degree by the superiority or inferiority of a group. Second, in the geography investigation, whatever differences there may have been in teaching skill were rotated out, since each procedure was taught by each teacher. Third, whatever differences there may have been in unit or test difficulty were likewise rotated out as a factor affecting the results of the study.[7]

---

[1] At the beginning of the school year the pupils were ranked according to I.Q. and all of the pupils with odd numbered rankings formed one class and all with even numbered rankings formed the second. We have thus, essentially, matched groups on the basis of I.Q.

[2] W. A. McCall, *How to Experiment in Education*, pp. 32–33. New York: Macmillan Co., 1923.

[3] Wallace Atwood and Helen Thomas, *Nations Beyond the Seas*. Boston: Ginn and Company, 1930.

[4] C. J. Pieper and Wilbur Beauchamp, *Every Day Problems in Science*. Chicago: Scott Foresman and Co., 1933.

[5] For copies of the tests used see C. D. Jayne, "The Integrated versus the Non-integrated Use of Classroom Moving Pictures." Appendix B. Ph. M. Thesis. University of Wisconsin, 1936.

[6] *Ibid.*, Table I, p. 14 for an analysis of the test items to determine the number covered by the book alone, by the film alone, and by both the book and film.

[7] W. A. McCall, *op. cit.*, pp. 32–33.

TABLE I

SUMMARY OF PROCEDURE

General Science Units	Same General Science Teacher Class "A"	 Class "B"
Hot Air Heating	Non-integrated (Procedure I)	Integrated (Procedure II)
Simple Machines	Integrated (Procedure II)	Non-integrated (Procedure I)
Geography Units	Class "A" Teacher No. 1	Class "B" Teacher No. 2
Austria	Non-integrated (Procedure I)	Integrated (Procedure II)
Hungary	Integrated (Procedure II)	Non-Integrated (Procedure I)

A definite time schedule was set up for the teachers to follow so as to make the time element constant for each of the units of work. Tables II and III show in detail the time schedules as they were originally drawn up for the geography and general science units respectively.

Detailed directions regarding the procedures for teaching the units of work were prepared for the use of the teachers so as to

TABLE II

TIME SCHEDULE FOR GEOGRAPHY

Unit: Austria

Class "A" Procedure I			Class "B" Procedure II		
Date:	Activity	Minutes	Date:	Activity	Minutes
1–24	Unit Pre-test	15	2–17	Unit Pre-test	15
1–30	Film presentation	20	2–17	Film presentation	42
1–31	Film test	10	2–18	Film test	10
2–13	Book and oral work	57	2–18	Book and oral work	47
2–14	Book and oral work	27	2–19	Book and oral work	15
2–14	Unit test	15	2–19	Unit test	15
	Total time	144		Total time	144

Unit: Hungary

Class "A" Procedure II			Class "B" Procedure I		
Date:	Activity	Minutes	Date:	Activity	Minutes
2–10	Unit pre-test	15	1–30	Unit pre-test	15
2–19	Film presentation	42	2– 5	Film presentation	20
2–20	Film test	10	2– 6	Film test	10
2–20	Book and oral work	47	2–20	Book and oral work	57
2–21	Book and oral work	15	2–21	Book and oral work	27
2–21	Unit test	15	2–21	Unit test	15
	Total time	144		Total time	144

TABLE III

TIME SCHEDULE FOR GENERAL SCIENCE

Unit: Hot Air Heating

Class "A" Procedure I			Class "B" Procedure II		
Date:	Activity	Minutes	Date:	Activity	Minutes
1–17	Unit pre-test	15	2– 5	Unit pre-test	15
1–23	Film presentation	20	2– 5	Film presentation	42
1–24	Film test	10	2– 6	Film test	10
2– 6	Book and oral work	57	2– 7	Book and oral work	47
2– 7	Book and oral work	42	2– 7	Book and oral work	30
2– 7	Unit test	15	2– 7	Unit test	15
	Total time	159		Total time	159

Unit: Simple Machines

Class "A" Procedure II			Class "B" Procedure I		
Date:	Activity	Minutes	Date:	Activity	Minutes
3–16	Unit pre-test	15	2–25	Unit pre-test	15
3–16	Film presentation	42	3– 3	Film presentation	20
3–17	Film test	10	3– 4	Film test	10
3–17	Book and oral work	47	3–16	Book and oral work	57
3–18	Book and oral work	30	3–17	Book and oral work	42
3–18	Unit test	15	3–17	Unit test	15
	Total time	159		Total time	159

standardize the general procedures used from class to class.[8] Directions for both procedures are summarized below.

Procedure I, which represents the non-integrated use of films, was carried out as follows:

1. The unit pre-test was administered. This included the film pre-test.
2. One week after the pre-tests the proper film was presented. There was no discussion of the unit in general or of the film itself before its presentation to the class. The discussion following the presentation was limited to answering questions asked by the pupils. Only 20 minutes were allowed for presentation and discussion.
3. The film test was given the day following the presentation of the film.
4. Two weeks after the film presentation, the formal study of the unit covered by the film was started. This formal study consisted of the reading and discussion of the text book material. In the geography classes this was supplemented by some additional material presented by the teacher or by oral reading from a reference book. The teachers made what use they could of scenes recalled from the moving pictures pertaining to the unit which had been seen two weeks before.
5. At the end of the formal study the final unit test was given. This was the same test that was used as a pre-test.

Procedure II, which represents the integrated use of films, was carried out as follows:

1. The unit pre-test was given. This included the film pre-test.
2. Immediately after the pre-test the film was presented to the class without discussion. After seeing the film there was a discussion of it and questions were raised. The film was then shown a second time to clear up points not understood. During this second showing the film was sometimes stopped while a point was discussed.
3. The day after seeing the film the final film test was given. This was the same test as had been used as the film pre-test.

[8] See C. D. Jayne, *op. cit.*, Appendix A, for copy of directions given to the teachers.

4. This test was followed immediately by the formal study of the unit which was done by the same general method as in Procedure I. It will be noticed, however, by reference to Tables II and III that the time allowance for formal study was shorter in Procedure II because of the longer time which had been devoted to the presentation of the film. The total time devoted to each procedure was the same.
5. At the end of the formal study the final unit test was given. This was the same test as had been used for the unit pre-test.

III. Analysis of the Statistical Results

This study, as has been pointed out before, consists of two independent investigations of the relative effectiveness of the integrated and the non-integrated methods of using educational moving pictures in the classroom. In analyzing the data from these investigations it seems well to consider first some of the general factors conditioning the results of both investigations; second, the data from the geography investigations; and third, the data from the general science investigations.

Turning our attention first to some of the general factors conditioning the results of both investigations, it should be noted that Procedure I was favored over the typical non-integrated method of using films by the pre-test situation, and by the position of the film lesson in relation to the formal study of the unit. Each of these advantages will be discussed in some detail.

It should be remembered that the pre-tests were given to the groups using Procedure I three weeks before they were given to the groups using Procedure II. In those cases where the class work covered during that period had a close relationship to the experimental unit upon which the pre-test was based it is probable that the pupils gained some information applicable to the experimental unit. The geography classes were studying other countries of Central Europe just prior to starting the units on Austria and Hungary and that this aided them in the unit pre-tests is suggested by the fact that the pre-test scores were higher when taken with Procedure II, as is indicated in Table V. This situation will be seen to give a definite advantage to Procedure I in the geography study, since all the information relative to Austria and Hungary gained in the three weeks preceding the start of the unit proper would be credited to Procedure I, but would not be credited to Procedure II, because in the latter instance gain is measured from the actual start of the unit when the pre-test was given. The general science classes were covering material less closely connected with the experimental units during the period between the pre-test (Procedure I) and the start of the experimental unit, and the pre-tests indicate but a slight carry-over, as may be seen from Table VIII.

Having considered the advantage accruing to Procedure I from the pre-test situation, attention will now be called to the fact that Procedure I represents practically the best possible use of the non-integrated method, so far as placement of the film in relation to the unit of work to which it pertains is concerned. Thus, in Procedure I the time interval between the seeing of the film and the starting of the formal work on the unit to which it pertained was only two weeks. That this was a most advantageous position will be realized when one considers that in the ordinary non-integrated use of the film this time interval will vary from a few days up to practically nine months, and that the average will probably be four or five months.

Another advantage growing out of the placement of the film lesson is that in Procedure I the film always preceded the study of the unit. In the ordinary non-integrated use of the film the picture would come as often after the unit as before it. Studies which have been made show that films used before a unit of work are more effective than those which follow the unit.

It should also be noted that in Procedure I there were no intervening pictures between the showing of a film and the study of its related unit. In the typical use of the non-integrated method there are generally intervening pictures. It seems probable that when this happens the children's memory of the picture will be less vivid when the formal study of the unit is started and that the amount of carry over from the film to the unit would therefore be less.

The preceding paragraphs make two points which need to be kept in mind. First, the pre-test situation in this study gave an advantage to Procedure I. Second, Procedure I represented exceedingly favorable conditions for the non-integrated use of films so far as the position of the film in relation to the

teaching unit is concerned. What this study actually does is to compare a typical method of integrating films with a unit of work, with superior conditions for the non-integrated use of films.

The general makeup and validity of the tests used have been discussed, but attention now needs to be turned to their reliability, as indicated by data from both pre-tests and final tests. The reliability of the various tests, as calculated by the chance half method and then stepped up by use of the Spearman-Brown prophecy formula, is shown in Table IV.

to meet the requirements for a valid single group experiment.

Thus the "A" Class in geography studied the unit on Austria using Procedure I and the Unit on Hungary using Procedure II, as is shown in Table I. If it can be shown that the unit on Austria is of comparable difficulty with that of Hungary, we have all the requirements for a single group experiment, the difference between Procedure I and Procedure II being the only variable. By use of the pre-test scores from all 56 cases over both units of work it is possible to show the comparative difficulty of the units. The mean pre-

TABLE IV

RELIABILITY OF TESTS

Unit	Unit Pre-test	Unit Final test	Film Pre-test	Film Final test
Austria	.56	.87	.46	.62
Hungary	.75	.77	.64	.78
Hot Air Heating	.68	.78	.89	.63
Simple Machines	.69	.75	.68	.75

Having pointed out some of the general factors which entered into the outcomes of the investigations, an analysis will now be made of the data from the geography investigation. It will be recalled that two units of work, Austria and Hungary, were used with two equated sixth grade classes. The analysis is based only upon data from students who were present for all of the work in both units. Due to extremely cold weather at the time the study was made, a number of the children were absent so that of the 79 who participated in the study data can be used from only 56 cases.

The individual scores and the gains made by each of the 56 pupils considered were tabulated and from these data Tables V, VI, and VII were prepared.

Due to projection difficulties which necessitated the stopping of the projector and the rewinding of the film during Procedure II the results from the "B" Class in geography were invalidated. However, the results from the "A" Class were in no way influenced by this accident and it can be shown that the experimental situation with this group was such as

test score from the unit on Austria was 30.8 with a P. E. of .57 and from the unit on Hungary was 29.4 with a P. E. of .42. The test results therefore indicate that the units are of practically the same level of difficulty. Thus the results from a single class may be considered as a valid single group experiment, and conclusions from the geography investigation are based on data from the "A" group.

Turning now to the results from the "A" Class, it will be found that a mean gain of 17.5 was made on the unit test by Procedure I and a gain of 21.1 by Procedure II, as shown by Table VI, there being a difference of 3.6 in favor of Procedure II. This difference has a critical ratio of 2.72, which in terms of chances is 99.7 in 100. On a percentage basis the unit gains were 21% greater from Procedure II than from Procedure I. Considering the unit gains made by individual pupils, it was found that 59% made a larger gain with Procedure II, 37% made a larger gain with Procedure I, and 4% did equally well with each procedure.

The results from the tests over the film material alone were decisively in favor of

TABLE V

SUMMARY OF DATA FROM GEOGRAPHY CLASSES, BY PROCEDURES, BASED ON 56 MATCHED CASES

Kind of test	Unit Tests		Film Tests	
Procedure	I	II	I	II
Mean pre-test score	29.04	31.1	21.1	22.0
Mean final test score	47.32	50.5	30.8	32.6
Mean unit gain	18.28	19.4	9.7	10.6
σ Gain	6.7	5.9	4.35	4.55
σ M Gain	.89	.78	.58	.60

TABLE VI

SUMMARY OF DATA BY CLASSES FROM GEOGRAPHY UNIT TESTS

Class	A	A	B	B
Unit	Hungary	Austria	Hungary	Austria
Procedure	II	I	I	II
Number of pupils	27	27	29	29
Mean unit pre-test score	30.6	29.9	28.3	31.6
Mean unit final test score	51.7	47.4	47.3	49.4
Mean unit gain	21.1	17.5	19.0	17.8
σ Unit gains	6.4	6.4	7.1	5.4
σ M Unit gains	1.23	1.23	1.3	1.0

TABLE VII

SUMMARY OF DATA BY CLASSES FROM GEOGRAPHY FILM TESTS

Class	A	A	B	B
Unit	Hungary	Austria	Hungary	Austria
Procedure	II	I	I	II
Number of pupils	27	27	29	29
Mean film pre-test score	20.4	22.7	19.6	23.5
Mean film final test score	33.2	31.2	30.5	32.0
Mean film gain	12.8	8.5	14.9	8.5
σ Film gain	4.86	4.68	4.10	4.59
σ M Film gain	.93	.90	.76	.85

Procedure II. Thus the film test following Procedure I showed a mean gain of 8.5 as compared with a mean gain of 12.8 on the film test following Procedure II, as is shown by Table VII. This difference is statistically significant, having a critical ratio of 3.98. The film gain by Procedure II was 51% greater than by Procedure I.

To summarize the results from the geography investigation, it may be said that Procedure II produced greater informational gains both in the unit tests and the film tests; that these gains were statistically significant; and that quantitatively they were 21% greater on the unit tests and 51% greater on the film tests.

The general science investigation was carried through without any extraneous factors entering in so far as could be determined. The number of pupils present during the teaching of both units was greatly reduced by extreme weather conditions which caused a very large number of absences. Of the 80 pupils who participated in the study only 30 had perfect attendance during both units. The statistical analysis is based upon these 30 cases.

Taking for consideration first the results from the unit tests it will be noted that the pupils made a mean score of 7.7 by Procedure I, and 11.1 by Procedure II (Table VIII), a difference of 3.4 in favor of Procedure II. This difference is statistically significant, since it has a critical ratio of 3.2. In terms of percent, Procedure II produced an increase in gain of 44% over Procedure I. Seventy percent of the pupils did better with Procedure II than with Procedure I, 23% did better with Procedure I, and the remainder did equally well with each procedure.

The difference between the two procedures, as measured by the film tests, was even greater, as is shown by Table VIII. Thus, the mean gain from the film test in Procedure I was 2.5 and the mean gain by Procedure II was 6.9, a difference of 4.4 in favor of Procedure II. This difference is statistically significant, having a critical ratio of 4.73. Expressed in terms of percent, Procedure II caused a gain of 176% more than Procedure I. The data on individual gains show that 76% of the pupils made a better film score by Procedure II, 16% did better with Procedure I, and the remainder did equally well with each.

Turning now to the separate classes we find that the "A" Class made a mean unit gain of 9.25 by Procedure I and a mean unit gain of 13 by Procedure II, as may be seen by reference to Table IX. This is a difference of 3.75 in favor of Procedure II and represents an increase of 40% over the gain made by Procedure I. In the film tests the same class made a mean gain of 4.25 by Procedure I and 6.41 by Procedure II (Table X). This difference of 2.16 represents a gain by the use of Procedure II of 50% over the gain made by Procedure I.

The "B" Class in the unit tests made a mean gain of 6.66 by Procedure I and 9.77 by Procedure II (Table IX). The difference in favor of Procedure II is 3.11 and this, in terms of gain over Procedure I, equals 46%. In the film tests the "B" Class made a mean gain of 1.44 by Procedure I, but a mean gain of 7.22 by Procedure II (Table X). This difference of 5.78 represents a gain by Procedure II more than 400% greater than by Procedure I.

The results from the general science investigation may be summarized by saying that Procedure II produced a unit gain 44% greater than Procedure I, and a film gain 176% greater than Procedure I. Both of these differences are statistically significant. Not only were the results from the investigation as a whole in favor of Procedure II, but each class and each unit of work show that more information was imparted by Procedure II.

IV. Conclusions, Limitations and Suggestions for Further Study

Conclusions drawn from this investigation can of course be considered applicable only to situations similar in every way to those of the present experimental study. There are many possible ways of integrating films with units of work, and there are limitless variations in the time interval between the showing of the film, and the study of the unit to which it pertains, in the non-integrated method of using films. The experiment was set up in such a fashion that Procedure II was considered to be a good method, but not necessarily the best, of using the film by the integrated method. Procedure I was considered to represent a favorable situation, far better than the average, for the incidental non-integrated use of films, because of the short time interval between the presentation of the film and the study of the unit, and because in each case the film preceded the unit.

TABLE VIII

SUMMARY OF DATA BY PROCEDURES FROM GENERAL SCIENCE UNIT AND FILM TESTS, BASED ON 30 MATCHED CASES

Kind of test	Unit Tests		Film Tests	
Procedure	I	II	I	II
Mean pre-test score	28.06	28.70	18.76	19.16
Mean final test score	35.76	39.76	21.33	26.06
Mean Gain	7.7	11.1	2.5	6.9
σ Gain	6.1	4.0	3.66	3.27
σ M Gain	1.11	.72	.66	.57

TABLE IX

SUMMARY OF DATA BY CLASSES FROM GENERAL SCIENCE UNIT TESTS

Class	A	A	B	B
Unit	Hot Air Heating	Simple Machines	Simple Machines	Hot Air Heating
Procedure	I	II	I	II
Number of pupils	12	12	18	18
Mean unit pre-test scores	34.75	23.0	23.61	32.5
Mean unit final test scores	44.0	35.0	30.27	42.27
Mean unit gain	9.25	13.0	6.66	9.77
σ Gain	6.3	3.8	6.0	4.1
σ M Gain	1.82	1.09	1.41	.96

TABLE X

SUMMARY OF DATA BY CLASSES FROM GENERAL SCIENCE FILM TESTS

Class	A	A	B	B
Unit	Hot Air Heating	Simple Machines	Simple Machines	Hot Air Heating
Procedure	I	II	I	II
Number of pupils	12	12	18	18
Mean fim pre-test scores	23.08	15.56	15.88	21.5
Mean film final test scores	27.33	22.08	17.33	28.72
Mean film gain	4.25	6.41	1.44	7.22
σ Gain	3.9	3.5	3.5	.3
σ M Gain	1.1	1.04	.82	.7

In brief the conclusions from these investigations are:

1. Children learn much more from a film when it is integrated with a unit of work, as was done in Procedure II, than when non-integrated as was done in Procedure I.

2. Children learn more from a unit of work when a film is integrated with it, as was done in Procedure II, than from a unit when a film is used but not integrated with it as in Procedure I.

3. The difference in the informational learning produced from the study of a unit, in favor of Procedure II, is statistically significant, and in this study was 21% greater in geography and 44% greater in general science. These differences are so great that they should be very important factors in determining which method is to be used.

While great care has been taken in the collection of data, the investigation is not without limitations. Probably the most obvious limitation is the small number of pupils from whom data could be used. It is to be regretted that the number of cases was reduced by weather conditions and by projection difficulties to less than 50% of those starting the study. Thus the conclusions in the geography study are based on one class of 27 pupils and the conclusions from the general science study upon two classes totaling only 30 cases. The smallness of the pupil population represented must make all conclusions tentative.

Other limitations grow out of facts already presented regarding the advantages probably accruing to Procedure I out of the pre-test situation and the very favorable position of the film in relation to the formal study of the unit. Another factor, not measured at all in this study, is the effect of breaking into a unit of work to project a film unrelated to the unit, as was done in Procedure I. It seems quite possible that the learning from a unit thus interrupted by an uncorrelated film might be affected, and if so, it is a factor which should be taken into consideration.

The above considerations suggest that additional research is needed on this problem. A larger population should be sampled in order to verify the findings of this study. The investigation needs to be extended so as to measure the gains on the unit which is being taught when the film is shown in Procedure I. Investigations should also be made to determine the effect of showing the film after the study of the unit, instead of before. It would also be profitable to vary Procedure I, so that the film would be used as in Procedure II, and the only variable would be the fact that in one procedure the film was shown outside the learning unit, and in the other during the learning unit. A greater variety of films should be included, and they should be analyzed in some detail to determine more definitely the effect of the character of the film upon the effectiveness of each method of presentation. Studies should also be made to determine the relative value of the integrated and non-integrated methods of showing films upon other types of learning than the informational. Thus it would be worthwhile to measure the relative effectiveness of the two procedures in producing changes in attitudes, understandings, etc.

It will be only after extended studies of the types indicated above have been made, that a complete picture can be presented of the relative advantages and disadvantages of the integrated and non-integrated methods of using educational films. This study has taken a common procedure with non-integrated films and has compared it with a common practice of integrating films with units of work. The conclusion, based upon the data collected, is that with the procedures used, the gain in information produced by the integrated method is significantly greater than that produced by the non-integrated method.

THE RELATIVE EFFICIENCY OF INSTRUCTION IN LARGE AND SMALL CLASSES ON THREE ABILITY LEVELS[1]

LACEY ARNOLD EASTBURN
Director Research and Guidance
Phoenix Union High School and Junior College, Arizona

The investigation here reported concerned itself with the degree to which students attained important objectives of instruction when taught in large and small classes homogeneously grouped on three ability levels. The problem may be more clearly stated in the form of questions: (1) Can the size of class be increased beyond the traditional limit of thirty in eleventh grade American history and English without decreasing the effectiveness of the instruction, as measured by objective tests, when the classes are homogeneously grouped? (2) Is the relationship which exists between the size of class and the effectiveness of instruction the same for homogeneous classes on three different levels of mental ability? (3) What effect does the size of class have upon the development of desirable attitudes in the students on different ability levels? (4) What are the reactions of teachers to large homogeneous classes?

Organization of the Study.—The experiment was carried on in the Phoenix Union High School during the school year 1933–34. During the year three large and six small classes in eleventh grade English were organized as follows: on each of the three ability levels one large class of approximately sixty students was paired with two small classes of approximately thirty pupils each. The same 360 students were then regrouped forming three large and six small classes in American history, with one large class of sixty and two small classes of thirty on each of the three ability levels. The regrouping was done in such a way that the sixty students who formed the large class in English formed the two small classes in American history, and likewise the sixty students forming the large class in American history constituted the two small classes in English.

The two small classes and the large one with which they were paired on the same ability level were taught the same subject by the same instructor. During the year of this experiment each instructor was assigned one class of thirty students in addition to the one large and two small classes included in the experiment. The pupil-period load was unchanged, but the customary five classes were replaced by one large and three small ones. The teaching load was increased somewhat due to a more extensive testing program than had been customary in previous years.

Method of Pairing.—The large and small classes were paired student to student on the basis of six criteria. The criteria used in pairing were: (1) raw scores on the Terman Group Test of Mental Ability; (2) raw scores on the Columbia Research Bureau American History Test; (3) raw scores on the Columbia Research Bureau English Test; (4) the total grade points earned in the ninth and tenth grades; (5) chronological age; and (6) sex.

The classes as a whole into which the students were grouped were remarkably well equated, the difference between the means on all criteria being well within the probable error of the difference. This is shown by Table I which gives the means of the pairing data of the different groups.

Methods of Teaching.—The six teachers participating in the experiment were allowed free latitude in their selection of the teaching procedure which they used, except that they were requested to employ the same general type of teaching procedure in both the large and the small classes. The result was that each teacher used the technique which she had been accustomed to employ in medium-sized classes.

[1] Abstract of doctoral dissertation, 1936, on file in the Stanford University Library.

TABLE I

MEANS AND SIGMAS OF THE SCORES ON THE TERMAN, ENGLISH AND AMERICAN HISTORY TESTS, PREVIOUS GRADE POINTS, AND CHRONOLOGICAL AGE OF THE PAIRED STUDENTS IN THE LARGE AND SMALL CLASSES ON THE DIFFERENT ABILITY LEVELS

Ability Level	Class	Measure	Scores on tests			Grade points first two years	Chronological age in months
			Terman mental tests	C. R. B. English test	C. R. B. American History test		
Upper	Large	Mean	164.75	136.71	52.48	42.33	191.42
	Small	Mean	164.82	135.29	53.10	42.49	191.02
		Diff.	—0.07	1.42	—0.52	—0.16	—0.40
Upper	Large	Sigma	14.77	26.05	16.66	10.78	8.29
	Small	Sigma	15.10	23.71	17.55	11.06	8.71
		Diff.	—0.33	2.34	—0.89	—0.28	—0.42
Middle	Large	Mean	132.20	109.85	39.55	35.56	194.03
	Small	Mean	132.22	108.93	39.05	35.27	195.10
		Diff.	—0.02	0.92	0.50	0.29	—1.07
Middle	Large	Sigma	8.02	17.45	13.12	10.00	10.94
	Small	Sigma	8.05	19.84	13.95	10.22	9.96
		Diff.	—0.03	—2.39	—0.83	—0.22	0.98
Lower	Large	Mean	97.72	90.36	31.98	29.02	198.28
	Small	Mean	97.91	89.41	31.57	29.60	198.48
		Diff.	—0.19	0.95	0.41	—0.58	—0.20
Lower	Large	Sigma	17.22	22.82	11.87	8.28	10.48
	Small	Sigma	17.12	22.58	12.01	9.76	11.39
		Diff.	0.10	0.24	—0.14	—1.48	—0.91

Courses of Study.—The course of study in American history for the classes on the upper and middle ability levels was the usual course given in the eleventh or twelfth grade of the senior high school. The instructor in American history of the classes on the lower level placed a great deal of emphasis on group and individual projects, with somewhat less stress on the strictly academic material emphasized on the other levels.

The course of study in English on the upper and middle ability levels consisted largely of English literature during the first semester and of American literature during the second. In both semesters some attention was given to composition, spelling, and vocabulary building. On the lower ability level much emphasis was placed on such current periodical literature as seemed to be within the comprehension of such children, with less emphasis placed on strictly academic work.

MEASUREMENT OF ACHIEVEMENT

Achievement was measured by means of objective tests. Both standardized and new type tests were utilized.

Standardized Tests. — The standardized tests used in measuring achievement were:

a. Columbia Research Bureau American History Test[2]

b. Iowa General Information Test in American History[3]

c. Columbia Research Bureau English Test[4]

[2] Harry J. Corman, Thomas N. Barrows, and Ben D. Wood. *Columbia Research Bureau American History Test.* Yonkers: World Book Company, 1927.

[3] M. H. Degraff, G. M. Ruch, and H. A. Greene, *Iowa General Information Test in American History.* Iowa City: University of Iowa. 1927.

[4] Harrison R. Steeves, Allan Abbott, and Ben D. Wood, *Columbia Research Bureau English Test.* Yonkers: World Book Company, 1925.

d. Iowa Placement Test in English Training[5]

e. Iowa High School Content Examination —Literature Section[6]

All of these standardized tests were given at the beginning of the year and a comparable form repeated at the close. The differences of the final scores over the initial scores were used as a measure of achievement.

New Type Objective Tests.—New type tests were constructed by the teachers covering the work of the semester as a whole, as well as each unit separately.

a. Each teacher constructed comprehensive semester examinations of 200 or more test items. In each case these examinations were given at the beginning of the semester and repeated at the close. The scores made on these tests were treated in the same manner as were the scores on the standardized tests.

b. *Unit Tests.*—Each teacher worked out an objective test of one hundred or more test items covering the material presented in each unit. The number of unit tests given varied according to the number of units into which the courses were divided, which number ranged from nine with some teachers to twelve with other.

Results of Achievement Tests.—The mean of the differences in scores of the students and the probable error ratios were computed according to "Students" method as described by Ezekiel.[7] The chances in one hundred of a real difference existing between the classes were taken from the table given by Garrett.[8]

The scores made on the unit tests, and the gains made on all other tests repeated were reduced to comparable series by means of the method given by Hull and described by Garrett.[9] The sum of the scores made on the different tests was treated by "Students" method as discussed above. The means of the differences and the probable error ratios obtained from the total of all test scores for the combined small classes and the large class with which they were mated are given in Table II. The data presented in this table show that the differences in the total of all scores obtained on the achievement tests for the year in American history and English on the upper ability level and American history on the middle ability level were in favor of the small classes. The probable error ratios in these three cases varied from 0.75 to 1.47. The obtained differences may thus be ascribed to the operation of chance factors.

The differences obtained between the large and small classes in English on the middle ability level gave a probable error ratio of 4.81. The probability that such a difference may be due to the operation of chance factors is exceedingly small, so that we may regard it as statistically significant.

The large classes on the lower ability level in American history and in English exceeded their mated small classes. The differences were not large enough to be statistically significant.

Comparison of Teachers' Marks Given in the Experimental Classes.—It is generally admitted that objective tests, regardless of how well they may be constructed, do not measure all the desirable outcomes of instruction. Certainly they do not measure all the outcomes which the majority of teachers attempt to take into consideration when they compute semester grades. In spite of all the well known limitations of teachers' marks, it was thought that a comparison of the marks given in the different classes might give a measure of the efficiency of instruction in large and small classes additional to that obtained by objective tests.

For the purpose of comparing the grades earned by the large and small experimental classes in American history and English, grade-point values were assigned as follows: a grade of "1" was assigned a grade-point value of four; "2", three; "3", two; "4", one; and "F" a value of zero. Table III gives the mean grade points, the mean of the differences, and the probable error ratios between the grade points of the students in the small classes and those of their mates in the large class on the different ability levels and subjects. Since the highest probable error ratio obtained is 2.86, none of the differences

[5] M. F. Carpenter, G. D. Stoddard under the direction of C. E. Seashore, and G. M. Ruch, *Iowa Placement Examination, English Training.* Iowa City: University of Iowa, 1925.

[6] G. M. Ruch, C. L. Huffaker, F. B. Knight, and W. Koerth, *Iowa High School Content Examination.* Iowa City: University of Iowa, 1924.

[7] Mordecai Ezekiel, "Students' Method for Measuring the Significance of a Difference Between Matched Groups," *The Journal of Educational Psychology,* XXIII (September, 1932), 446–50.

[8] Henry E. Garrett, *Statistics in Psychology and Education,* p. 271. New York: Longmans, Green and Company, 1926. This technique, described by Garrett, is questioned by some statisticians because it assumes that the obtained difference is a true difference and succeeding obtained scores would distribute normally about it.

[9] *Ibid.,* p. 282.

TABLE II

SUMMARY OF THE MEANS OF THE DIFFERENCES AND THE PROBABLE ERROR RATIOS BETWEEN THE TOTAL SCORES ON ALL ACHIEVEMENT TESTS OF THE PUPILS IN THE SMALL CLASSES AND THAT OF THEIR MATES IN THE LARGE CLASSES

Class	Ability Level	Size of Class	Mean of Diff.	$\frac{\text{Mean}_d}{PE_d}$	Chances in 100
American History	Upper Upper	Large Small	0.98*	0.75	69
English	Upper Upper	Large Small	1.41	1.47	84
American History	Middle Middle	Large Small	1.92	1.31	81
English	Middle Middle	Large Small	—5.08	—4.81	99.99
American History	Lower Lower	Large Small	—2.40	—1.83	89
English	Lower	Large	—0.96	—0.91	73

*The differences were obtained by subtracting the score of the student in the large class from that of his mate in the small class.

TABLE III

SUMMARY OF THE MEAN GRADE POINTS OF THE DIFFERENT CLASSES, THE MEAN OF THE DIFFERENCES, AND THE PROBABLE ERROR RATIOS BETWEEN THE GRADE POINTS, (GIVEN ON BOTH SEMESTERS), OF THE STUDENTS IN THE SMALL CLASSES AND THOSE OF THEIR MATES IN THE LARGE CLASS ON THE DIFFERENT ABILITY LEVELS AND SUBJECTS

Class	Ability level	Size of class	Mean grade points	Mean of diff.	$\frac{\text{Mean}_d}{PE_d}$	Chances in 100
American History	Upper Upper	Large Small	5.10 5.58	0.48	2.86	97
English	Upper Upper	Large Small	4.90 5.31	0.42	2.50	95
American History	Middle Middle	Large Small	4.12 4.82	0.70	2.57	96
English	Middle Middle	Large Small	4.44 4.18	—0.26	—1.12	77
American History	Lower Lower	Large Small	3.59 3.37	—0.22	—0.95	74
English	Lower Lower	Large Small	3.69 3.67	—0.02	—0.01	50

can be said to be highly significant statistically.

Comparison of the data of this table with those of Table I shows that the small classes in American history on the upper and middle ability levels slightly exceeded their mates in both the achievement tests and semester grades received. The large classes in English on the middle and lower ability levels and American history on the lower level slightly excelled their mates in the small classes.

Student Attitudes Toward Large Classes.—In an attempt to get an objective measure of the student attitudes developed in large classes, the Hand-Carley "Student Reaction Form"[10] was given to the 360 students of the experimental classes. This student questionnaire was developed by Hand and Carley of Stanford University for the purpose, to quote the authors, "of determining the validity of various hypotheses concerning the more or less intangible but important outcomes associated with large class instruction. . . . This instrument yields what, for lack of a better name, its authors call an 'attitude-to-be-desired score' ".[11]

By means of this instrument it was possible to compare the attitudes of students in large classes with those of students in small classes, with the same teacher in the same subject. Table IV gives the summary of the means of the differences and the probable error ratios between the attitude-to-be-desired scores of the students in the small classes and those of their mates in the large class on the Student Reaction Form on the three levels of mental ability. From the data presented in this table it will be seen that the students of the large class in American history on the upper ability level and the large class in English on the lower ability level made higher attitude-to-be-desired scores than did their mates in the small classes. In the other four cases the small classes excelled their mates. None of the differences proved to be highly significant. One might conclude, then, that the size of class has little effect in determining the development of desirable student attitudes.

Conclusions

The following conclusions are drawn from the findings of the present class-size investigation briefly reported in these pages.

1. Students in the eleventh grade in American history and English on the upper ability level and in American history on the middle ability level did slightly better in classes of thirty than in classes of sixty, when measured

Table IV

Summary of the Means of the Differences and the Probable Error Ratios Between the Attitude-to-be-Desired Scores of the Pupils in the Small Classes and Those of Their Mates in the Large Classes

Class	Ability level	Size of class	Mean of diff.	$\frac{\text{Mean}_d}{PE_d}$	Chances in 100
American History	Upper	Large	—0.71	—1.03	76
	Upper	Small			
English	Upper	Large			
	Upper	Small	6.46	3.34	99
American History	Middle	Large			
	Middle	Small	2.35	1.08	77
English	Middle	Large			
	Middle	Small	0.10	.05	51
American History	Lower	Large			
	Lower	Small	6.42	3.05	98
English	Lower	Large	—4.42	—2.07	92
	Lower	Small			

[10] Harold C. Hand and Verna A. Carley, Student Reaction Form (privately printed, not published for distribution).
[11] Verna A. Carley and Harold C. Hand, "Students' Attitude Toward Large and Small Classes," Stanford University, unpublished manuscript.

by objective tests, but the difference was no greater than could be accounted for by the operation of chance factors.

2. Eleventh grade students of the middle ability level in English did better when measured by objective tests in a class of sixty than in classes of thirty. The difference was greater than could be accounted for by chance factors. These findings might not hold true for all teachers, but they do show that some teachers can handle large classes of middle ability students in English more effectively than they handle small classes.

3. Eleventh grade students of the lower ability level in American history and in English did slightly better in classes of sixty than in classes of thirty, though the small difference obtained could be accounted for by chance factors. This casts some doubt upon the often repeated statement that students of the lower ability level must be taught in small classes if the greatest efficiency is to be secured.

4. In five of the six experiments, students not accustomed to large classes continued to improve in their relative accomplishment for some period of time after being placed in large classes.

5. Class size does not appear to be a vital factor in determining the development of the forty-three attitudes measured by the Hand-Carley "Student Reaction Form". The majority of students agree: (1) that a greater variety of opinions are expressed in large than in small classes, and (2) that large classes foster the development of a greater degree of self-dependence.

6. The replies to the questionnaire given to the teachers (a complete description of which is included in the present writer's thesis on file in the Stanford University Library) indicate that there is little unanimity of opinion among teachers, who have had experience with large and small classes, on the different arguments advanced in favor of, or opposed to, large classes.

AN EXPERIMENTAL STUDY OF THE EFFECT OF EDUCATIONAL GUIDANCE ON A SELECTED GROUP OF HIGH SCHOOL SOPHOMORES

T. Ernest Newland
Bucknell University, Lewisburg, Pennsylvania
and
Willard E. Ackley
High School Principal, Donora, Pennsylvania

In spite of the fact that much has been written about and claimed for the practice of guidance in connection with education, surprisingly little experimental evidence has been obtained on it. Studies such as reported by Becker[1] and Brumbaugh[2] have indicated that groups on which some type of guidance was practiced felt some effect, but no evidence was presented to show what happened to a similar group not so influenced. One of the most objective studies found was a parallel-group experiment reported by Lund in which the process of educational guidance consisted of "remedial work" in the form of interviews with the experimental group, the negligible results being qualified by a claim that the personal interview technique might be shown to be better by a different attack on the problem, and also by the plea that the "personal interview in the secondary schools is an activity that merits further scientific investigation."[3] That there has been an unfortunate lack of experimental investigation on problems of this sort has been pointed out by Ayer[4] and others. It is also unfortunate that studies in this field have been directed all to frequently toward the prevention of failure, usually among lower capacity groups, rather than at the problem of the discovery of practices and materials which can be used by those concerned with educational guidance in an attempt to cause students of superior capacity to achieve more in harmony with their capacity to achieve.[5]

This study is somewhat like the others in that it was based upon failing or potentially failing students, but it does contribute findings based upon controlled experimentation. Stated specifically, the problem investigated was: What was the effect of a program of educational guidance upon a group of academic problem pupils in a sophomore class in high school, particularly as regards (a) the earning of quality credits, (b) the failing of subjects, and (c) the withdrawal from school?

Procedure

On the basis of the educational history of the freshman class, there was compiled a list of 52 students—40 boys and 12 girls—who seemed to the principal and teachers most likely to be cases of potential failure during the sophomore year. On the basis of sex and average mental age, as obtained on the Terman Group Test of Mental Ability and the Otis Self-Administering Test of Mental Ability, an attempt was made to obtain parallel groups of matched pairs. While the average mental ages were practically identical, the experimental group was slightly more heterogeneous than the control group—the mean M. A. for the experimental group being 179 months, with a sigma of 17.42, and the mean M. A. for the control group being 180 months, with a sigma of 15.12. Three of the 26 pairs accounted largely for this difference, the discrepancies in these cases being 14, 15, and 10 months.

[1] Elsa G. Becker, "First Annual Report of the Guidance Department of the Samuel J. Tilden High School," *High Points*, (November, 1933), 19.
[2] H. E. Brumbaugh, "Reducing Failures by a Plan of Alternative Dismissal, Holding Failing Pupils for Additional Instruction," *Journal of Educational Research*, XXIII (March, 1931), 240–45.
[3] S. E. Torsten Lund, "The Personal Interview in High School Guidance," *School Review*, XXXIX (March, 1931), 196–207.
[4] Fred C. Ayer, "School Progress," *Review of Educational Research*, III (June, 1933), 194–200.
[5] It is true that we have the A. Q. and devices such as that published this year by Inor for R. D. Allen, but these are only beginnings.

When the educational guidance, which was essentially remedial work, began, each pupil in the experimental group was required to report to the principal's office on Monday morning before the first class period and to secure a weekly report sheet which he took to his classes throughout that day. Each teacher of each subject taken by the pupil wrote in the name of the subject taken, the comment "satisfactory" or "unsatisfactory" to describe the nature of the work done the preceding week, and signed his name. At the end of the day the pupils brought the sheets to the office, where the results were recorded on a master record sheet. Every pupil who received "unsatisfactory" reports in one or more subjects was called in for a discussion of his situation. If the pupil was not able to state the reason for his unsatisfactory report, he was required to ask for a conference with the teacher to determine the cause. He then reported to the investigator for further consideration of the matter and suggestions were made to the pupil in the light of what was known of his background and capacities. If the work was marked "unsatisfactory" the following week, the pupil was again called in for conference, after which a report of the condition was mailed to his home.

The status of each pupil in the experimental group was studied from several angles in order to be able to work more efficiently with him. In addition to an initial interview with the pupil (from four to eleven interviews were held with each one in the experimental group during the time of the study from December to the following June), the homes of these youngsters were visited and the cooperation of the parents, none of whom happened to be foreigners, was encouraged. A careful study was made of the school and health records, which were augmented by results of other tests and inventories, in addition to the intelligence test results mentioned earlier: a 150-item home-made objective diagnostic test in English, the Iowa Silent Reading Test, the Bernreuter Personality Inventory, a 27-item home-made study habits inventory, and the Haggerty-Olson-Wickman Behavior Rating Scale (Schedule B), the latter of which was checked by each experimental pupil's roll-call and subject-matter teachers. Once rapport was established between the pupil and the interviewer, the student was shown whatever evidence was desirable in trying to get him to feel, in a cooperative sense, that the school expected him to reach that achievement level for which he was equipped. With no attempt at apparently intending to do so, all those working with the pupils endeavored to make them feel that the staff had a real, personal interest in the pupils' welfare.

While this program carried with it a certain element of contagion in that it spread to some of the other pupils in the school, thus perhaps causing a few in the control group to profit by the motivation supplied especially by means of the weekly reports, it can be safely said that the remedial and motivation work affected the experimental group in a more intentional and consistent manner than it may have the control group.

Results

The significance of the amount of academic superiority of the experimental group over the control group can not well be determined upon any one basis of comparison, but must be seen in the light of at least two aspects of the situation. While the experimental group as sophomores showed greater improvement over their freshman records in terms of quality credits earned than did the control group, the experimental group also had fewer students leave school. As freshmen, the pupils in the experimental group earned an average of .81 quality credits per subject taken (A=4, B=3, C=2, D=1, E=0), and as sophomores they earned an average of 1.37 quality credits per subject, a gain of .56 quality credit. Comparable data on the control group show that group gain to be .36 quality credit, which would seem to indicate a slight advantage in favor of the experimental group (columns 1 and 2, Table I).

But this advantage in favor of the experimental group is spuriously reduced by the fact that the control group suffered greater mortality and especially since the primary reason for this withdrawal from school of the six control pupils was academic failure. One student in the experimental group withdrew from school, but this was for financial reasons, his work at that time being of passing quality. Correcting the data to include these cases, it will be seen, in columns 3 and 4 of Table I, that the advantage in favor of the experimental group becomes greater.

A comparison of the number of subjects failed by the two groups, as presented in

TABLE I

AVERAGE NUMBER OF QUALITY CREDITS EARNED PER SUBJECT TAKEN

	Including performance only of students who stayed in school		Including performance of students who withdrew	
	(1)	(2)	(3)	(4)
	Experimental N=25	Control N=20	Experimental N=26	Control N=26
Average as freshmen	.81	.89	.84	.89
Average as sophomores	1.37	1.25	1.41	.97
Gain	.56	.36	.57	.08
Difference	.20		.49	

Table II, shows an advantage in favor of the experimental group, even when only the records of those remaining in school are used. The contrast becomes still greater when the records of those withdrawing are included. The decrease in the number of subjects in which the 20 remaining control subjects failed (from 10 to 6 out of a total of 80 subjects taken—see column 2) is relatively not as great as it may seem at first. The control group really failed in 13 per cent of the courses taken, as freshmen, and failed in 7 per cent of the courses taken as sophomores. In the other columns, the numbers of failures can be read, roughly, as percentages.

CONCLUSIONS

On the basis of these data, it seems safe to conclude that when educational guidance, largely in the form of remedial and motivational work, was administered to an experimental group, and the academic performances of this group compared with those of a similarly constituted group: (a) a better quality of school work, as measured by marks, was done by the experimental than by the control group; (b) proportionately fewer failures occurred in the "guided" group; and (c) while six out of the 26 in the control group dropped out of school primarily on account of failing marks, none of the experimental group withdrew for this reason.

TABLE II

NUMBER OF SUBJECTS IN WHICH FAILURE OCCURRED*

	Including only students who stayed in school		Including students who withdrew from school	
	(1)	(2)	(3)	(4)
	Experimental	Control	Experimental	Control
As freshmen	19	10	20	10
As sophomores	3	6	3	21

*Numbers are the same as in Table I.

FREE COMPOSITIONS AS AN AID TO THE TEACHER IN CHILD GUIDANCE*

GENEVIEVE M. LEARY
State Normal School, New Haven, Connecticut

THE PROBLEM

The purpose of the study was to determine whether or not free compositions of children are valid tools for use by the teacher or student teacher in child guidance. When the composition data were validated the writer was concerned with the grade, school, and sex interests of children, as revealed in the free writing, as secondary purposes. Other revelations, which might be termed subjective, since they were not verified in the criterion, were recorded. Another purpose of the writer was to study the interests children revealed in free drawings, and to compare these interests with those revealed in their free compositions, in order to find out what, if anything, free drawings added to the teacher's knowledge of a child in addition to the composition data.

Explanation of the Term, "Free" Compositions.—"Free" compositions are those written by children on topics of their own choice without any guidance or interference on the part of teachers in so far as content or theme is concerned. The papers used in this study were left untouched by a teacher. The spelling was not corrected, nor was the sentence structure or choice of words criticized with respect to the specific pupil. The teacher was free to read them for the purpose of getting ideas as to what needs children had on the technical side of written composition so that pertinent instruction might be given on these needs.

Because of the realization that genuine free compositions can be expected only in a classroom situation where the teacher is temperamentally capable of "keeping hands off" of child products, and wherein children feel secure because of the desirable relationships existing among classmates, between child and teacher, and teacher and child, the writer made a selection of twelve class groups for study out of a group of seventeen middle or intermediate grades available, after a careful study of the entire situation.

THE STUDY SITUATION

The study was carried on in New Haven, Connecticut, in twelve classrooms in three laboratory schools affiliated with the New Haven State Normal School. The child groups in each of the three schools vary as far as the average economic and social backgrounds are concerned. One school environment is a favorable residential section, another school is in a semi-business and residential section and is somewhat less favorable, and the third school presents a rather deprived environment. The study situation, in short, presents a sampling of three varying environmental backgrounds in a Connecticut city.

There were four class groups on each of the three grade levels, grades four, five, and six, used in the study. Four hundred fifty children (146 in the sixth grade, 150 in the fifth grade, and 154 in the fourth grade) ranging in chronological age from 8 years and 3 months to 13 years and 3 months, were involved in the study which was carried on from October 1, 1934 through the second week of June, 1935.

THE METHOD

Collection of Data—The procedure followed in getting the free compositions was to set aside two periods a week when children would know that they would have the opportunity of doing this kind of writing. This was understood as a standing procedure. At the beginning of the period help was given in spelling when the child acted as if his expression were blocked by a realization of the inadequacy of his spelling. Each child put his paper into his individual folder and these were preserved throughout the year. In June the writer asked permission of each grade

* Abstract of a dissertation presented for the Doctor of Philosophy degree at Yale University.

group to borrow their folders for her own private use, and this permission was granted. Thus the composition data came into the hands of the writer.

Means for Validation of Data.—During a period of six weeks during the month of May, 1935, and the early part of June, 1935, the writer, assisted by seven student teachers on the fourth-year level of the teachers college program of studies, interviewed all the children included in the study. The writer presented the interview sheet to a competent psychiatrist, who approved it if given before any analysis of the individual's compositions was made, so as to eliminate any "ploughing" or searching inquiry to attempt to bring forth things which would tend to explain the revelations of the papers. This direction was followed. The interview sheet contained thirty questions which were asked of the child unless he answered them inadvertently while answering other questions. A standard approach was given in each instance. Each interview took at least fifteen minutes. There were two questions which were always asked in the same order by the interviewer. These were: Number 1, asked first, What do you like to do when you are not in school? and Number 2, asked last, In what are you especially interested? Phrases or words were used in recording whenever possible.

The writer trained the seven interviewers by interviewing individual children while two or three interviewers sat behind a partial wall provided by a cloak room listening to the interviewer and the child. Due to the cooperation of the laboratory school teachers one child at a time came out to an interviewer, the selection of the child being left to chance, that is, any child who could conveniently leave his work at the time came out to be interviewed when another child returned to the room.

As the interview sheets were turned in, the name on every tenth one was recorded for a home visitation to be made by the writer. These visits were made and the findings relative to parental judgment of child interests and general behavior of 45 children were recorded.

Another means of validation sought was a report on each child given by the teacher in June, 1935, after working and living in the classroom with the child during the school year. The inquiry sheet called for the name of the child, a statement of his interests, his awareness of what was going on about him or in the world, and for any remarks which the teacher might care to make.

To summarize briefly, an interview with each child, the teacher's report on each child, and the home visitation of every tenth child interviewed, were used as means for the validation of the composition data.

Analysis of Data.—After analyzing a sampling of data to get enough encouragement to attempt the study, the writer obtained the services of eight people to read and to analyze the papers according to what they felt the children revealed about themselves in their writing. The title or topic of every paper for each child was recorded with a word or two regarding it noted after the title. When all the papers for a given child had been treated in this way, the reader wrote a sketch of the child as she knew him through the compositions, that is, his interests. A random sampling of the composition data of one half of the 450 cases available for study was taken for analysis. This sampling provided 9,004 free compositions.

The Criterion

A brief sketch of the individual child as revealed by the interview, the teacher's report, and the home visitation report when available, was made. This composite picture became the criterion for the validation of the material revealed through an analysis of the child's compositions. When a goodly number of criterion sheets and corresponding composition analysis sheets had accumulated, the composition sheets, without any names of children appearing upon them, were paired by two people, independently.

The ratio of consistency between the specific items in the composition data sketch and those in the criterion sketch was then determined. The paired cases were taken and the items on each sheet were checked by two persons, other than the writers of the sketches. The number of items on the criterion sheet became the denominator of the fraction expressing the ratio, and the number of items revealed by the sketch based on the composition data which were also in the criterion became the numerator.

$$\frac{\text{Number of Items Common to the Composition and Criterion Sketch}}{\text{Total Number of Items in Criterion Sketch}} = \text{Ratio of Consistency}$$

In order to check the findings still further, the sampling of paired criterion sketches and composition analysis sketches was divided in half, that is, half of each of the twelve grade groups, and the ratio of consistency between the criteria and the composition data was determined for each of the two halves of the original paired sampling.

The writer of the study made a third sheet or sketch, in addition to the criterion and composition analysis sketches made by other persons. Upon this were recorded child likes, dislikes, and revelations such as his attitudes toward adults, and some character traits. These sheets, without the name of the child, but lettered in code, were submitted to the teacher who had worked with the children for an entire year, to see whether or not she could recognize them. After reading them, she labelled them with the appropriate name, according to her judgment, and returned them to the writer.

Sample Third Sheet—Reader's Judgment (Identification letter—d)

Very philosophical turn of mind—serious—mature outlook—interested in explaining, seems to clarify own thinking as explanation develops—extremely observant—longs for a dog—muses about a wide range of subjects, such as own early childhood mischief as related by mother—also muses on behavior of parents, especially the father, at whom the child "takes digs" at times—interested in people's problems—rather solitary as far as child companions are concerned—goes about much with parents, but doesn't enjoy it especially.

The Use of Free Drawings

A collection was made of 1,936 free drawings done by the 225 children participating in this study over a period of two months. The drawings were made under the same conditions described for the writing of the free compositions, that is, the child was a free agent. He could draw whatever he pleased in whatever medium he cared to use: pencil, crayon, or water color. Although the mode of expression was a different one from that of writing, the writer wished to see if the drawings added to the picture presented by the composition material.

The drawings were studied to find the answers to the following questions:

1. What per cent of the drawing interests are also found in the criterion and the composition analysis sketches of each child?

$$\frac{\text{Number of Drawing Interests Found in Criterion and Composition Analysis}}{\text{Total Number of Drawing Interests}} = x \text{ per cent}$$

2. What do the drawing interests add to the compositions when there are some not found in the composition sketch as expressed by a ratio?

$$\frac{\text{Number of Drawing Interests not in the Composition Sketch}}{\text{Total Number of Composition Interests}} = x \text{ per cent}$$

The Findings

Free compositions written in school situations where children feel secure, and where teacher dictation of content is absent, constitute valid data in revealing the interests of children for the following reasons:

1. Composition sketches made on the basis of interests revealed in free compositions were made by persons other than those who wrote the corresponding criterion sketches. These criterion sketches were paired accurately with the composition analysis sketches by two persons who had written neither sketch in 208 of the 225 cases or in 92.4 per cent of the half sampling of cases.

2. Sixty-two per cent of the items revealed in the composition analysis sketches were also found in the criterion sketch. When the half sampling of the total number of cases was again halved according to class groups, 61 per cent of the items in the composition analysis sketches were also found in the criterion sketches in the first half, and 63 per cent were the same in the second half.

The children studied ranged in age from 8 years and 3 months to 13 years and 3 months, and presented a wide variation in interests. The major classification of interests revealed is the following: personal experiences, including play, lessons outside of school, camping, and creative expression; hobbies and clubs; pets; nature; religious affiliations; reading of fiction and non-fiction; sports, both individual and group; science; radio; current events; travel; school activi-

ties; and a miscellaneous group of scattering interests. The highest five classifications of the total nineteen for the entire child group are as follows:

	Per Cent
1. Reading	21.06
2. Sports	19.25
3. Hobbies and clubs	10.82
4. Pets	6.72
5. Play	6.42

The lowest five classifications are as follows:

	Per Cent
1. Camping	.49
2. Religious affiliations	.71
3. Earning money	.75
4. Radio	1.63
5. School activities	1.70

The economic, social, and physical environmental differences and influences are quite marked in the interests revealed by children living and attending school in three varying environmental situations.

It is likely that the things children revealed about themselves in the free writing, which were not revealed by their teachers, their parents, or by their own interviews, are of importance and worthy of some consideration by teachers interested in promoting desirable growth in children, since such revelations are found among other items that were verified. Among these are likes and dislikes, fears, dreams, social and moral standards, and traits of all kinds.

Teachers identified 46.75 per cent of the sketches based on these additional unproved items gleaned from the compositions of children with whom they had worked for one year.

Free drawings made under the same conditions as the free writing, that is, without dictation of content, and without group or teacher criticism after the product was completed, add, on the average, 22.5 per cent to the interest picture of the child found in his compositions. To the eye of a trained analyst, they may reveal items, equally important in child study and guidance to those described above in connection with free compositions, which the writer did not verify. They are worthy of consideration as still another tool to be used in guidance.

Suggestions for Obtaining, Preserving, and Using Free Composition Data

It is suggested that teachers set aside time when pupils may write freely, and that they have children keep these writings in sturdy folders over a period of a year, together with any other free writing they do. These folders may be used by teachers and student teachers in laboratory schools throughout the year as a readily accessible means of getting acquainted with the children, in addition to observation and actual contacts with them in the activities of the school. They may also be kept by the teacher after the children have gone into the next grade, for use the next year by the teacher with whom the children are then working. When free drawings have been added to the folders by the children, the tool may become even more valuable.

It would be well for student teachers to have some experience in studying free compositions and free drawings of children during their period of preparation, under the guidance of a psychiatrist, to make the future teacher aware of how she may use these data in gaining a better understanding of the children with whom she works.

THE RELATION OF READING COMPREHENSION AND RETENTION TO HEARING COMPREHENSION AND RETENTION[1]

WILLIAM E. YOUNG
Syracuse University

I. THE PROBLEM

This investigation was undertaken to provide data pertinent to the answering of the following major issue: What is the relation of comprehension and retention in reading to comprehension and retention in hearing? It was thought that data pertinent to this question would also prove helpful in answering a second: To what extent is reading ability a result of good language comprehension?

The problem of this study is of fundamental importance, for its solution should imply the correct reading procedures both for first teaching and remedial work; for example: Should the pupil who is deficient in silent reading ability be given thought-getting training and language exercises or should he be given drills to facilitate the various phases of eye movements? Are the mechanical factors of the reading activity such as the duration of eye fixation pauses, the number of eye fixation pauses a line, the return sweep of the eyes from the end of one line to the beginning of the next, and the like, to be attacked or taught directly, or are they in the large to be acquired as a result of language and reading growth?

As can be seen from a review of studies in this field (particularly Yoakam's study),[2] it is known that children get very little from a single silent reading. The small returns from a silent reading may be due primarily to low language comprehension on the part of the pupils or primarily to their lack of such proficiencies as are peculiar to the act of reading the printed page. Some of these proficiencies may be connected with the mechanical or motor skills. No one yet has given a clear picture of the abilities needed in reading over and above those comprehended in language efficiency.

How much do pupils who get very little from an unassisted silent reading obtain from an oral teacher presentation? This question the experimenter determined to answer.

Even though the returns from a teacher oral presentation might prove to be small, quite as small as those found in other studies for the silent reading, definite conclusions would be dependent upon the scores being comparable. Hence, it was decided to conduct again the experiment of finding the effect of a silent reading and to conduct this experiment with the same cases and the same reading materials at approximately the same time to ascertain the comparative effects of different modes of presenting various types of content.

Should one provide a language situation which removes the mechanical or motor factors of silent reading and obtain results which emphasize low language comprehension as the main obstacle to overcome, still one would not know but what there were individuals whose reading disability was due to the mechanical or motor factors. In a group of one hundred, for example, ten might do well after an oral presentation but very poorly after a silent reading, while another ten of the same group might do quite the reverse. The result might be that the means of the two sets of scores would be approximately equal. Therefore, it was decided to work out correlation coefficients between the scores on orally presented material and the scores on material read silently. These correlation coefficients would

[1] Abstract of a thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the College of Education of the State University of Iowa, 1930.

[2] G. A. Yoakam, "The Effect of a Single Reading," *Twentieth Yearbook* of the National Society for the Study of Education, Part II, Report of the Society's Committee on Silent Reading, pp. 90–102. Bloomington, Illinois: Public School Publishing Company, 1921.

furnish an index of the degree of correspondence or "go-togetherness" between the two abilities.

The study was also planned so as to provide quartile comparisons. Those pupils who did the poorest after a silent reading, those who fell in the lowest quarter, were examined to see where they placed after an oral teacher presentation. This analysis was done also for the pupils of the lowest quarter after the oral presentation. It was done again in the same double fashion for all pupils in the highest quarter.

Finally, it was decided to find out the grade improvement in comprehension through silent reading and that in comprehension through hearing. Buswell[3] has reported a plateau in the development of the eye motor habits in the intermediate grades. He believes that as a result reading development is arrested in this period. If this condition be true, one would expect hearing comprehension to progress more markedly in the intermediate grades than reading comprehension, held back, as the latter is reported to be, by an arrested development of the motor factors.

Very briefly the problem of the thesis was: What is the relation between reading comprehension and hearing comprehension?

To solve this problem the following subproblems were set up:

1. What is the effect of an oral presentation?
2. What is the comparative effect of different modes of presenting various types of content?
3. What is the degree of correspondence or "go-togetherness" between reading comprehension and hearing comprehension?
4. Do the children who do poorly in comprehending through reading (or hearing) do poorly in comprehending through hearing (or reading)?
5. Do the children who do extremely well (comparatively) in comprehending through reading (or hearing) do equally well in comprehending through hearing (or reading)?
6. What is the comparative status of reading comprehension and hearing comprehension at the various intermediate grade levels?

II. The Procedure

The pupils used in this experiment were taken from the intermediate grades, four, five, and six, of six school systems, of which five were in Iowa and one in Texas. About two thousand pupils took part in the experiment.

The reading materials were fifteen in number of four different types: three descriptions of industrial processes, four nature science units (units on wild animal life), four hero stories from American history, and four narrative poems. The selections varied in length from about three hundred and fifty words (the descriptions of industrial processes) to about eight hundred words (the hero stories).

These selections were presented in four different ways, each selection being presented merely one way to the same classroom, but each classroom (with only a few exceptions) experiencing at least three modes of presentation and three selections from each of the four types of reading materials. The four modes of presentation were: (1) Mode A, the teacher read aloud to the pupils; (2) Mode B, the teacher read aloud to the pupils while they read the selection silently; (3) Mode C, the pupils read the selection once silently at their own individual rate; and (4) Mode D, the pupils read the selection silently for the same amount of time assigned for the oral reading by the teacher.

Immediately after the presentation of each selection the teacher tested the pupils over it. One test of thirty items was used. The first twenty or so items were questions calling for short definite answers. These items formed a recall test. The last ten or so items were multiple choice units of three responses each. Whereas the recall items usually dealt with specific bits of knowledge, the multiple choice units were concerned with general impressions, deductions, etc. The tests of the nature science units were administered again one month later to secure a measure of delayed recall.

A special experiment was conducted with the pupils of one of the Iowa towns, Iowa 1. With the groups in this Iowa town, each test was not given at all until one month's time had elapsed from the presentation of the corresponding reading unit. This special exper-

[3] G. T. Buswell, *Fundamental Reading Habits: A Study of Their Development*, p. 57. Supplementary Educational Monographs, No. 21. Chicago: University of Chicago, 1922.

TABLE I

PER CENT SCORES OF COMBINED TOWNS ON TESTS

Grade	Nature Science			Poetry			Hero Stories			Industrial Processes		
	Mode			Mode			Mode			Mode		
	A*	C**	D†	A	C	D	A	C	D	A	C	D
4	33	25	29	23	20	22	27	23	24	21	18	18
5	41	35	41	36	32	31	40	35	37	32	24	32
6	46	41	44	37	31	34	49	44	.....	40	37	39

* Teacher oral presentation.
** A single silent reading.
† Silent reading with same time intervals as for oral presentation.

iment was carried out in order to get a measure of delayed recall without the stimulus and the learning factor of previous testing.

The tests were proved highly reliable. The coefficients of reliability in grade five for composite tests were from .87 ± .018 to .92 ± .012. The tests were also probably quite valid, correlating .54 ± .037 in grade four and .60 ± .036 in grade five with the Thorndike–McCall test, and correlating .67 ± .035 in grade six with the Gates Advanced tests.

Since there was a large number of cases involved and since any two groups which were compared were composed of approximately an equal number of children from the same town, and since the cross-checking technique was used (with the pupils of Mode A in one comparison as pupils of Mode B in another, Mode C in a third, and Mode D in a fourth), it was not thought necessary to equate the pupils of the different groups. However, this assumption itself was tested and pupils of groups compared were equated on the basis of their scores on the Gates Advanced Reading Tests. The results, as Tables XI and XII show, from such equated groups and from the groups not so equated were in essential harmony.

III. THE FINDINGS

The findings showed that children get very little from an oral presentation (21 to 33%—varying according to type of selection—in grade four) as measured by immediate testing. Children have retained an even smaller amount after one month's expiration, a further loss of 43% having taken place. If the children are not tested initially until one month has passed, the percentage of retention ranges from 7 to 20, varying according to type of selection. Tables I, II, and III show these findings.

TABLE II

PER CENT SCORES OF COMBINED TOWNS ON TESTS (OVER NATURE SCIENCE) REPEATED AFTER ONE MONTH

Grade	Mode		
	A	C	D
4	20	17	17
5	31	27	28
6	32	32	34

Children do get, however, more from a teacher oral presentation than from a silent reading by themselves. Tables IV and V give supporting data. The difference was greatest in grade four (where it was almost uniformly statistically significant) and almost absent in grade six. The method of teacher oral presentation plus simultaneous silent reading by the pupils was roughly equally as effective as the teacher oral presentation alone.

Even the few tables of results presented here are quite clear in showing that children

TABLE III

PER CENT SCORES OF IOWA 1 ON TESTS DELAYED FOR ONE MONTH

Grade	Nature Science				Poetry				Hero Stories				Industrial Processes			
	Mode				Mode				Mode				Mode			
	A	B*	C	D	A	B	C	D	A	B	C	D	A	B	C	D
4	19	15	13	17	20	16	16	14	11	10	9	11	7	....	4	5
5	23	24	18	20	24	26	21	20	17	17	17	17	9	....	8	10
6	24	24	22	24	28	25	20	22	18	20	14	19	10	....	9	12

* Teacher oral presentation plus simultaneous silent reading by pupils.

TABLE IV

DIFFERENCES IN TEST RESULTS BETWEEN THE AVERAGES OF MODE A (ORAL READING BY TEACHER) AND MODE C (SINGLE SILENT READING WITHOUT TIME LIMIT)

Type of Selection	Iowa Towns	Grade	No. of Cases	Diff. * A-C	P. E. Diff.	Diff. / P. E.
Nature Science	2; 3	4	227	2.384	.334	7.14
Nature Science	2; 3	5	235	1.804	.347	5.20
Poetry	2; 3	4	223	.535	.256	2.09
Poetry	4; 5	4	159	1.198	.346	3.46
Poetry	2; 3	5	244	1.096	.307	3.57
Poetry	4; 5	6	164	1.741	.444	3.92
Hero Stories	2; 5	4	140	1.333	.388	3.44
Hero Stories	2; 3	5	238	1.405	.369	3.81
Hero Stories	2; 3	6	241	1.575	.367	4.29
Industrial Processes	2; 3; 4; 5	4	395	.772	.149	5.18
Industrial Processes	2; 3; 4; 5	5	414	1.571	.168	9.35
Industrial Processes	2; 4; 5	6	251	.597	.269	2.22

*Positive difference in favor of Mode A.

improve in their ability to comprehend through reading throughout the intermediate grades quite as fast as, if not faster than, they improve in their ability to comprehend through hearing.

The coefficients of correlation obtained between the pupil results on tests after oral presentation and pupil results on tests after silent reading were uniformly high, practically as high as, and in some cases higher than, the pupil results on two different tests of silent reading. The Pearson product-moment coefficients of results on composite tests after oral presentation and on composite tests after silent reading were .80 ± .025 and above. Tables VI, VII, VIII, and IX give the coefficients of correlation obtained between different modes of presentation on individual tests. Table X gives the coefficients when results on tests over composite selections are used.

In general, children who do poorly in comprehending through reading do poorly in comprehending through hearing. No children were found to be in the highest quarter of one of these phases of language comprehension and in the lowest quarter of the other.

IV. EDUCATIONAL IMPLICATIONS

The net amount that children of the intermediate grades get from a single presentation (no matter what the mode of presentation)

TABLE V

DIFFERENCES IN TEST RESULTS BETWEEN THE AVERAGES OF MODE A AND MODE D (SILENT READING WITH TIME LIMIT EQUAL TO THAT OF THE ORAL READING)

Type of Selection	Iowa Towns	Grade	No. of Cases	Diff. * A-D	P. E. Diff.	Diff. / P. E.
Nature Science	2; 3	4	227	1.194	.342	3.49
Nature Science	2; 3	5	235	.157	.342	.46
Poetry	2; 3	4	223	—.278	.269	1.03
Poetry	4; 5	4	159	1.031	.354	2.91
Poetry	2; 3	5	244	1.594	.310	5.14
Poetry	4; 5	6	164	.796	.460	1.73
Hero Stories	2; 5	4	140	.942	.399	2.36
Hero Stories	2; 3	5	238	.836	.376	2.22
Industrial Processes	2; 3; 4; 5	4	395	.691	.154	4.49
Industrial Processes	2; 3; 4; 5	5	414	—.055	.185	.30
Industrial Processes	2; 4; 5	6	251	.394	.269	1.47

*Positive difference in favor of Mode A.

TABLE VI

COEFFICIENTS OF CORRELATION BETWEEN DIFFERENT MODES OF PRESENTING NATURE SCIENCE SELECTIONS

r	P. E.	No. of Cases	Grade	Town	Mode
.60	.043	99	4	Iowa 5 Iowa 2 Iowa 3	A* C**
.69	.035	100	4	Iowa 4 Iowa 3 Iowa 2	A C
.53	.036	182	4	Texas 1	A C
.57	.052	76	5	Iowa 3 Iowa 2 Iowa 5	A C
.66	.034	127	5	Iowa 4 Iowa 3 Iowa 2	A C
.63	.029	190	5	Texas 1	A B†
.53	.041	139	6	Iowa 3 Iowa 5 Iowa 2	A C
.61	.049	74	6	Iowa 3 Iowa 5	A C
.60	.042	110	6	Iowa 2 Iowa 4 Iowa 3	A C
.66	.037	107	6	Texas 1	A C

* Teacher oral presentation. ** A single silent reading.
† Teacher oral presentation plus simultaneous silent reading by pupils.

TABLE VII

COEFFICIENTS OF CORRELATION BETWEEN DIFFERENT MODES OF PRESENTING POETRY SELECTIONS

r	P. E.	No. of Cases	Grade	Town	Mode
.60	.049	77	4	Iowa 3, a, d Iowa 2	A* C**
.59	.051	77	4	Iowa 2 Iowa 3, a, d	A D†
.64	.034	136	4	Iowa 2 Iowa 5 Iowa 4 Iowa 3, a, d	C D
.55	.040	136	5	Iowa 2 Iowa 3 Iowa 4	A C
.64	.035	134	6	Iowa 2 Iowa 3 Iowa 5	A D
.54	.045	114	6	Iowa 2 Iowa 3 Iowa 4 Iowa 5	A C
.72	.036	80	6	Iowa 3	A C
.54	.053	81	6	Iowa 3	A D
.72	.037	78	6	Iowa 3	A C

* Teacher oral presentation.
** A single silent reading.
† Silent reading with same time intervals as for oral presentation.

was not determined exactly by this thesis. Although it can be said that the children of the fourth grade got about twenty per cent on a Mode C presentation (a single silent reading without time limit) of poems, it is not known but that the children would have made some score on the test even if the single presentation had not occurred. The experimenter found out from each teacher included in the study whether or not her class had read or studied the selection; and in all cases where the class or any part of it had read or studied the selection, the selection was not administered. But it cannot be maintained that all pupils who had experienced a selection were excluded from experiencing it again. However, the number of pupils and situations involved does enable one to conclude that although the exact effect of a mode of presentation cannot be stated, the comparative effect of the four modes should now be fairly well known.

Keeping in mind the fact that the scores on the immediate recall are probably an exaggeration of the effectiveness of a single presentation in that they are the result of what the pupils knew about the selection before it was presented plus what they got from the presentation itself, the reader should re-examine Tables I, II, and III. The fact to be noted is the falling off between the immediate test scores and those repeated a month later. The loss was at times well over forty per cent. When the tests were given one month after the presentation without the learning factor of the initial testing to enter in, the scores were

TABLE VIII

COEFFICIENTS OF CORRELATION BETWEEN DIFFERENT MODES OF PRESENTING HERO STORIES

r	P. E.	No. of Cases	Grade	Town	Mode
.59	.049	82	4	Iowa 2 Iowa 3 Iowa 5	A* C**
.67	.039	87	4	Iowa 2 Iowa 4 Iowa 5	A C
.55	.055	74	4	Iowa 2 Iowa 3 Iowa 5	A C
.60	.035	132	5	Iowa 2 Iowa 3 Iowa 4	A C
.70	.039	81	6	Iowa 3	A C
.70	.039	78	6	Iowa 2 Iowa 3 Iowa 5	A C
.63	.038	116	6	Iowa 2 Iowa 3 Iowa 4	A C

* Teacher oral presentation.
** A single silent reading.

so bunched at the end of the scale that nothing approaching "normality" was obtained. Out of nine presentations of descriptions of industrial processes, only one was more than ten per cent effective. It can be seen that if the factor of previous knowledge could be eliminated and a true picture of the effectiveness of each mode secured, it would point plainly to the conclusion that one presentation is not enough for learning. Probably the establishment of this principle in such an emphatic fashion with varying types of selections and varying modes of presentation is a more important contribution than the pointing out of one mode of presentation as more or less effective than another.

Another educational implication here is that the school must provide definitely for retention on the part of the child. Many writers on the subject of reading have failed to include memory or retention as one of the reading skills. Its essential importance as such cannot be denied. This study has also shown that the skill of retention needs to be developed in connection with oral learning quite as much as with silent reading.

Russell[4] established the seventh grade as the point where silent reading by the pupils "catches up" in effectiveness with the teacher oral presentation. Russell considered Grades V, VII, and IX in his investigation. Had he included Grade VI, he might have found that in work-type material, as exemplified in nature science materials and descriptions of industrial processes, silent reading by the pupils is probably equal in effectiveness to the teacher oral reading.

The point to be stressed here is not that Grade VI or Grade VII is the transition period but that the school can perhaps determine just when it will have the child's silent reading ability equal his ability to compre-

[4] R. D. Russell, "The Relative Effectiveness of Presenting Verbal Material Visually and Orally as Measured by the Amount of Recall," Unpublished Ph.D. Thesis, State University of Iowa, 1923.

TABLE IX

COEFFICIENTS OF CORRELATION BETWEEN DIFFERENT MODES OF PRESENTING DESCRIPTIONS OF INDUSTRIAL PROCESSES

r	P. E.	No. of Cases	Grade	Town	Mode
.52	.046	113	4	Iowa 2 Iowa 3 Iowa 4 Iowa 5	A* C**
.53	.046	110	4	Iowa 2 Iowa 3 Iowa 4 Iowa 5	A D†
.52	.046	113	4	Iowa 2 Iowa 3 Iowa 4 Iowa 5	C D
.53	.044	125	5	Iowa 2 Iowa 3 Iowa 4 Iowa 5	A D
.61	.035	153	5	Iowa 2 Iowa 3 Iowa 4 Iowa 5	A C
.56	.040	130	5	Iowa 2 Iowa 3 Iowa 4 Iowa 5	A C
.45	.048	125	5	Iowa 2 Iowa 3 Iowa 4 Iowa 5	C D
.75	.028	112	6	Iowa 2 Iowa 3 Iowa 4 Iowa 5	A C
.52	.050	97	6	Iowa 2 Iowa 3 Iowa 4 Iowa 5	A C

* Teacher oral presentation.
** A single silent reading.
† Silent reading with same time intervals as for oral presentation.

hend through hearing. The findings of this thesis would indicate that the school could readily establish this point in Grade V, provided educators deem such an accomplishment necessary and desirable.

There was no plateau discovered in silent reading growth in the intermediate grades to parallel the plateau reported by Buswell.[5] In fact, the findings of this thesis would indicate that improvement in silent reading compre-

[5] G. T. Buswell, *op. cit.*, p. 57.

TABLE X

COEFFICIENTS OF CORRELATION BETWEEN DIFFERENT MODES OF PRESENTING COMPOSITE SELECTIONS

r	P. E.	No. of Cases	Grade	Town	Mode
.78	.030	79	5	Iowa 2 Iowa 3	D† A*
.88	.019	72	5	Iowa 2 Iowa 3	D C**
.80	.025	101	5	Iowa 2 Iowa 3 Iowa 4	C A

* Teacher oral presentation.
** A single silent reading.
† Silent reading with same time intervals as for oral presentation.

TABLE XI

SCORES OF PUPILS IN EQUATED GROUPS (IOWA 2, IOWA 3, IOWA 4, AND IOWA 5) ON DESCRIPTIONS OF INDUSTRIAL PROCESSES, GRADE 4

No. of Pupils	Gates Reading Grade Averages	S. D. of Gates Reading Grades
104	4.865	1.266
94	4.867	1.268
97	4.887	1.283

Number of Pupils Experiencing

Selections	Mode A	Mode C	Mode D
L. P.	94	97	104
P.	104	94	97
T.	97	104	94
Av. Score	4.4576	3.7220	3.8475

$$\text{A av.—C av.} = .736 \pm .175 \frac{\text{Diff. (A-C)}}{\text{P. E. Diff.}} = 4.21$$

$$\text{A av.—D av.} = .610 \pm .181 \frac{\text{Diff. (A-D)}}{\text{P. E. Diff.}} = 3.37$$

hension is seemingly more accelerated than improvement in hearing comprehension in the intermediate grades.

The data of this thesis show that the relationship between reading comprehension and retention on the one hand and hearing comprehension and retention on the other is a very intimate and detailed one. The data suggest that the language or thought factor in silent reading constitutes much of the reading problem. Reading seems to be almost wholly a central as opposed to a peripheral activity.

It may well be that much of so-called reading disability is poor language comprehension. A method of teaching reading predicated on general language effectiveness (language and bodily activity of the thought-getting type) may be the solution to many of the problems in the teaching of reading. Such a method would give the child experiences of an eventful character, would have him express those experiences in both spoken and written words, and would then introduce him into the reading of his own and like experiences. It is not intended to suggest that this is a new method of teaching reading, but rather, that this study has apparently given it additional validation as contrasted with a method which seeks to focus attention and drill on the mechanical or motor phases of reading.

Certainly the data of this study should not be interpreted to mean that the school need not teach the child to read, provided it teaches him to hear. There are reading skills as distinguished from those of hearing. The abilities to use the index and the table of contents are examples. Possibly there are

TABLE XII

PUPILS' SCORES (IOWA 2, IOWA 3, IOWA 4, IOWA 5; GRADE 4) ON "HOW PINS ARE MADE" (APPLICATION OF LINDQUIST'S FORMULA FOR MATCHED GROUPS)

	No. of Pupils	Gates Reading Grade Averages	S. D. of Gates Reading Grades	Average Score on Pins
Mode A	104	4.865	1.266	5.740
Mode C	94	4.867	1.268	4.266
Mode D	97	4.887	1.283	4.722

A av.—C av.=1.474 $\pm$ .235 (formula for matched groups) $\frac{\text{Diff. (A-C)}}{\text{P. E. Diff.}} = 6.273$

A av.—D av.=1.018 $\pm$ .246 (formula for matched groups) $\frac{\text{Diff. (A-D)}}{\text{P. E. Diff.}} = 4.138$

skills connected with the factor of eye movements which should be taught. In first instruction and even more so in remedial work there is need for direct teaching of techniques in the use of the printed page. These certain definite skills peculiar to reading should be acquired but they should be acquired in thought-getting situations and through language or bodily activities functioning as meaningful responses.

The investigator believes that the data probably do mean that any thought-getting situation or any meaningful language activity does provide training of the best kind for reading and that any reading skill is perhaps best taught when placed in such a setting.

Furthermore, the findings do justify somewhat the implication that in any case of reading disability the first step should be to ascertain the language difficulties of the case. The knowledge that much of reading difficulty is probably language or meaning difficulty should be cheering in the sense that the average school is probably better able to cope with this aspect of reading than with the perhaps more technical phases of binocular fusion, eye fixation pauses, return sweeps of the eye, and the like.

AN ANALYTICAL STUDY OF THE READING HABITS OF DEAF CHILDREN*

TRUDA GOUGH LAGRONE

THE PROBLEM

The basic difference between deaf and hearing persons lies in their ability to use language. The interpretation of printed language through reading is one of the most complex phases of language usage; hence an analysis of the reading habits of deaf children offers opportunity for the study of one of the highest types of language activities.

A well-recognized method for analytical study of reading habits is that of making photographic records of eye-movements. The eye-movements of the reader indicate symptoms of three fundamental elements of the reading process. The average number of fixations per line is purported to measure the span of recognition for the printed material, and to indicate to what degree the reader has approached the higher order of perceptual habits necessary for maturity in reading. The average duration of the fixation pauses is believed to measure the reader's rate of recognition of the perceptual unit, and the average number of regressive movements per line is assumed to indicate the regularity or rhythmic progressive sequence of perceptions along the printed line.

The photographic technique is valuable, therefore, for throwing light on certain developmental phases of reading which cannot be ascertained through any other method of measurement. Photographic records of the eye-movements of a reader offer material for an analysis of the complex process of his reading into "its constituent processes, and reveal the development in the specific factors which go to make up reading ability."[1]

For several years investigators have been interested in studying various abilities and achievements of deaf children, and in studying the effect which their inability to acquire language in the normal way has upon their motor, associative, and perceptual learning; but no one has offered an analytical study of their habits of reading, revealed through photographic records of their eye-movements, as a method of investigating the effects of their language handicap upon the development of the reading process.

Photographic records of the eye-movements of hearing children show that reading is a process of rapid growth in its fundamental elements during the first four years of school experience, at the end of which time maturity is approximated; and that a high degree of correlation exists between the maturity of the subject in the fundamental eye-movement habits and his ability to comprehend printed material.[2]

In his study of the types and causes of failure in learning to read, Gray[3] lists fourteen significant ones, among which are poor auditory memory, inadequate speaking vocabulary, small meaning vocabulary, and speech defects. Obviously, deaf children suffer from these four deficiencies. What effect do these deficiencies have upon the deaf child's reading habits?

Investigations[4, 5, 6] show that for hearing children, ability in word perception correlates

---

* This paper is constructed from a portion of a thesis presented to the faculty of The University of Texas in 1936 in partial fulfillment of the requirements for the degree of Doctor of Philosophy. Professor C. T. Gray, of the Department of Educational Psychology, directed the study. For the sake of brevity, the initial survey of related literature has been omitted in this report.

[1] John A. O'Brien, *Reading: Its Psychology and Pedagogy*, p. 87. New York: The Century Company, 1926.

[2] Guy T. Buswell, *Fundamental Reading Habits: A Study of Their Development*, pp. 20–21. Supplementary Educational Monographs, No. 21. Chicago: University of Chicago, 1922.

[3] William S. Gray, *Remedial Cases in Reading: Their Diagnosis and Treatment*, p. 21. Supplementary Educational Monographs, No. 22. Chicago: University of Chicago, 1922.

[4] R. Pintner and D. G. Paterson, "A Measurement of the Language Ability of Deaf Children," *Psychological Review*, XXIII (November, 1916), 413–436.

[5] R. Pintner and D. G. Paterson, "The Ability of Deaf and Hearing Children to Follow Directions," *Pedagogical Seminary*, XXIII (December, 1916), 477–497.

[6] Arthur I. Gates, "Methods and Theories of Learning to Spell Tested by Studies of Deaf Children," *Journal of Educational Psychology*, XVII (1926), 289–300.

highly with ability to read, and that although deaf children surpass hearing children of the same reading age, both in spelling and word perception, they are markedly inferior in their ability to interpret printed material in context. Do the eye-movement records of deaf children throw any light on this lack of relationship between these apparently related factors? Do deaf children indicate this lack of interpretative ability by numerous regressive movements? Do deaf children fixate on words in rapid sequence, regardless of the degree of comprehension, with few regressions? Do deaf children ever reach a stage of maturity in language symbolism at which units instead of words are perceived? Is there a high degree of variability among the individuals on the different maturity levels of these elements, or does the level of maturity appear to be a function of school experience, and, therefore, to be correlated highly with school grade and years in school?

This investigation is an attempt to throw some light on these and other related problems with reference to the reading of deaf children. Stated more specifically, the purpose of this investigation of the reading habits of deaf children is: (1) to determine the presence or absence of developmental stages in three fundamental elements of the reading habit; (2) to study the degree of relationship between the fundamental elements in reading and other measurable factors related to the reading process; (3) to make an analytical study of individual performances in the reading of deaf children of different school grades and different mental age levels.

The Psychological Significance of the Problem

Theoretically, the hearing adult has a verbal substitute for every object in his world. These word substitutes become associated with various situations and experiences, and make possible the development of generalizations, abstractions, and recognitions of relationships which extend the individual's environment and increase the variability of his behavior into an almost unlimited refinement of categories.[7, 8]

In the development of a system of symbolic habits, audition normally plays a prominent role. The hearing person, through random articulation with the fixation of circular responses, imitation, and conditioning experiences,[9] learns the language symbols of his social group. By the manipulation of words, he virtually carries his world with him. Words function in calling out responsive behavior as truly as do the objects and situations for which they serve as substitutes.

Reading is one phase of the complex symbolic system of language. Utilizing a background of sound symbols already familiar to him, the hearing child is able to acquire the complex behavior involved in the reading process. Oral reading is recognized as a valuable experience to him in his early efforts in the reading complex, since the ear is already trained to hear the sentence he is reading. The young hearing reader, by associating the visual symbols with the sounds and kinaesthetic experiences of spoken words and sentence units, is able to comprehend the printed material.

The deaf child, however, has an entirely different route to travel in acquiring the mechanics of reading. He enters school with neither spoken nor written vocabulary. His concepts, if not vague, are at least associated with a set of symbols other than those he must use in reading; and therefore, the medium for associating these concepts with printed symbols is limited, if not wholly lacking.

A study of the development of the reading habits of deaf children is considered psychologically significant, therefore, for two reasons:

(1) It offers a means of studying a phase of language interpretation by subjects uninfluenced by auditory imagery, auditory memory, and auditory conditioning, and by whom the formation of general and abstract ideas or generalizing experience has been accomplished through more or less controlled and limited mediums.

(2) Such a study offers data for comparing the genetic development of deaf children with that of hearing children in fundamental elements of reading as shown by the number of fixations per line, the average duration of fixations, and the average number of regressive movements per line; hence, it may serve

[7] Albert P. Weiss, *A Theoretical Basis of Human Behavior*, pp. 307–327. Columbus, Ohio: R. G. Adams and Company, 1929.
[8] John B. Watson, *Behaviorism*, p. 187. New York: W. W. Norton and Company, Inc., 1925.
[9] Floyd H. Allport, *Social Psychology*, pp. 181–184. Boston: Houghton Mifflin Company, 1924.

to stimulate thought toward a more effective methodology in instructing deaf subjects in the art of reading.

Selection of Subjects

The purpose of the study makes it necessary to eliminate, so far as possible, all variables other than deafness.

The first factor necessary to hold constant is the age of onset of deafness. The results of the study of Pintner and Paterson[10] show that while the age of onset of deafness had little influence on the intelligence rating as measured by the Pintner Non-Language Test, it did have a decided influence on the educational rating; each year of hearing after the age of three or four increased language facility. In order to make sure that audition had affected his language acquisition little or none, no subject was chosen who had lost his hearing later than two years of age.

The second factor to hold constant is the percentage of residual hearing.[11] Pupils whose percentage of hearing is 35 per cent or less are unable to profit by the auricular training[12] offered in the school.

The third factor to be controlled is the method of instruction. Pintner and Paterson[13] found that the ability to profit by the oral method correlates highly with mental ability, and that on the average, oral pupils are less retarded than those taught by other methods. It was not possible to obtain for this study students in the higher grades who had not had contact with the sign and manual methods of communication, but subjects were chosen whose training in the actual classroom activities had been by the oral method throughout their school experience. Since the subjects of this experiment were chosen from a school in which all children are first offered the opportunity of the oral method, it is probable that whether the ability to profit by the oral method is a factor of general intelligence, a "g" factor, or a special talent, an "s" factor, it is held fairly constant.

Of the experimental group 71.2 per cent are girls, but this predominance is perhaps insignificant for three reasons. First, the purpose of the study does not involve a study of sex differences. Second, the selective factors of the experiment, and the retention of the photographic records based upon their clearness, make sex more or less a matter of chance. Third, regardless of the number of girls, the mean I.Q.'s of the two groups (girls and boys) are almost identical. As measured by the Pintner Non-Language Test, the mean I.Q. of the 27 boys is 89.99, and of the 67 girls, 87.68. The highest I.Q. among the girls is 144.79, and among the boys 131.78. The lowest I.Q. among the girls is 51.39, and among the boys 60.27.

By an examination of the records of all pupils enrolled in the oral department in the Texas School for the Deaf, Austin, Texas, the experimenter found the names of 113 pupils who had lost their hearing at two years or under, and whose residual hearing was 35 per cent or under. These pupils were selected as subjects upon these criteria, and photographic records were made of their eye-movements. Only 95 of the records were clear enough to be usable. When the comprehension scores were checked, it was found that one of the readers had failed to indicate comprehension on the test, making it necessary for his record to be discarded. The records of 94 readers were classified as follows:

Table I

Classification of the Subjects of the Experiment

Grade	Boys	Girls	Total
1	0	5	5
2	1	6	7
3	3	14	17
4	8	8	16
5	1	11	12
6	2	3	5
7	4	6	10
8	2	5	7
9	1	3	4
10	1	2	3
11	4	3	7
Col. Prep.	0	1	1
Total	27	67	94

Contact With Subjects

Another important factor for a successful experiment with deaf children is the matter

[10] R. Pintner and D. G. Paterson, "A Mental Survey of the Deaf," *Journal of Educational Psychology*, XIX (1928), 497.

[11] The term, residual hearing, is used to designate the amount of hearing the child possesses, whether his deafness is congenital or due to accident or disease. This percentage is obtained by measuring the percentage of deafness and subtracting that amount from 100 per cent.

[12] This training is made possible by the use of an amplifying system which provides individual ear sets for the pupils.

[13] R. Pintner and D. G. Paterson, "A Class Test with Deaf Children," *Journal of Educational Psychology*, VI (1915), 598.

of personal contact of the experimenter with them. The obtaining of photographic records of eye-movements requires a strict obedience to directions and complete confidence in the experimenter. The experimenter made bi-weekly visits to the Texas School for the Deaf over a period of ten weeks. Her appearance in the classrooms soon ceased to be anything out of the ordinary. This association gave the students opportunity to become accustomed to her. The teachers explained to the students that the experimenter was interested in their reading and would, before long, take a picture of their eyes while they read. As a consequence the subjects showed curiosity instead of fear, when placed before the camera for the reading experiment.

Tests Given

The photographing of the eye-movements while reading was the last step in obtaining data for this experiment. Although the study is concerned chiefly with eye-movements, yet certain test data were desired, and several days were spent in administering tests, not only to the subjects of this study, but to all the student body of the Texas School for the Deaf.[14] The administration of the testing program served not only as a means of obtaining test data, but also as another means of orienting the subjects to the experimental situation.

Scores on the Pintner Non-Language Test and Form W of the New Stanford Achievement Test were obtained for all the subjects; and scores on the Detroit Reading Test, the Detroit Word Recognition Test, and the Gates Primary Reading Test were obtained for all subjects in the first three grades.

Selection of Reading Material

Since one of the purposes of the experiment is to compare the development of the reading habits of deaf children with those of hearing children, it was thought best to employ the same reading selections that Buswell used with hearing subjects. This was possible for all subjects above the second grade. Buswell's selection for his readers of the first grade, however, could not be used with deaf children in the first and second grades because it contained verb forms and other grammatical complications with which the deaf children had had no acquaintance. For instance, the sentence, "He wanted the little dog to go home," involved for the deaf child an understanding of the use of the infinitive, and an understanding of the substantive clause used as the object of a transitive verb. The purpose of this experiment was, in so far as possible, to obtain samples of reading records, not records of struggles with unfamiliar grammatical constructions. It became necessary, therefore, to compose a reading selection within the limits of the vocabulary and grammatical comprehension of deaf children of the first and second grades. The teachers of the first grades in the School for the Deaf submitted to the experimenter several selections which they knew incorporated no unusual words or grammatical complications. The experimenter combined parts of these selections into three paragraphs, two for practice selections and one for the experimental selection.

After the experimental selection for the first two grades of deaf children was arranged, the next step was to determine the difficulty of the selection as compared with the selection for the first graders in the Buswell study. For this test, forty-two hearing children in grade 2A in the Palm Elementary School, Austin, Texas, were used. In order to eliminate, so far as possible, fatigue, practice effect, and other extraneous factors, the children were divided into two groups of twenty-one subjects each. The two selections were presented to the two groups in rotating fashion; that is, to one group the Buswell selection was presented first, and to the other group, the experimental selection was presented first.[15] Immediately following the reading of the selections, the subjects were given a series of questions based on the content of the reading selection. These questions were in mimeographed form and usually required a response of one word. The papers were scored, with the following results: Group I, to which the Buswell selection was presented first, made a mean score of 62 on the Buswell selection, and a mean score of 47 on the experimental selection. Group II, to which the experimental selection was presented first, made a mean score on the Buswell selection

[14] The writer acknowledges with appreciation the services of Dr. H. T. Manuel, Professor of Educational Psychology, The University of Texas, and of his advanced class in tests and measurements in the testing program.

[15] It was through the kindness of Miss Leigh Peck, a graduate student at The University of Texas, and Miss Frances Malone, a teacher in the Palm School, that this test was made.

of 67, and on the experimental selection a mean score of 46.

The scores of both groups were practically the same for each of the selections, and both groups scored decidedly higher on the Buswell selection. So far as the groups of hearing children are concerned, the mean scores indicate that the Buswell selection was easier, and in only a few cases were individual variations in favor of the experimental selection.

Practice Selections

As stated previously, the experimenter combined parts of material submitted by the first grade teachers in the School for the Deaf to obtain the practice selections for grades one and two.

For the practice selections for the older groups, use was made of paragraphs from Gray's New Standardized Oral Reading Check Test, Set II, No. 2, and the Pressey Fundamental Reading Habits Test, Grade 2, Form A.

Reading Selections

For First and Second Grades

Practice Selections

One Sunday a little girl and her brother went for a walk. They saw an old woman. They laughed at her. They were impolite. Their father scolded them.

A boy rode his bicycle on the sidewalk one day. He ran over a little dog. The boy carried the dog home. After a while the dog died. The little boy was sorry.

Experimental Selection

One Saturday afternoon a little boy and his dog went for a walk. They went to the park and played. After a while the dog cut its foot on a piece of glass. The boy carried the dog to the dog hospital. A woman looked at the dog's foot. She tied a cloth around it.

For Third Grade and Above

Practice Selections

Gray's New Standardized Oral Reading Check Test, Set II, No. 2, for grades 2 and 3.

There is an old stone house in my town. It is not far from my home. No other house is near it. There is a large yard and a wall around it. When it was new it was a very fine house. No one wants to live in it now. The front door is made of wood. It is now open. There is no glass in the windows. I can see into the house. It is dark and very dirty. There is nothing in it but an old white cat. She goes there every day to catch rats and mice.

Practice Selection—Pressey Fundamental Reading Habits Test, Grade 2, Form A.

A little stone house stood on a high rock. In the house lived a little boy and his father and mother, and his brother. There were other boys living in other stone houses nearby, and sometimes they came over and played with our little boy.

One day the little boy and his brother were playing alone. The older people were busy and no one was watching the houses. Then all of a sudden, there was a big noise. Indians with feathers in their hair ran out into the valley below the houses. The Indians had come to kill all the people.

Experimental Selection—same as used by Buswell.

One night Peter went to bed early. It was not dark. The bright moon shone in at the window. Peter could see everything in the room. All at once he heard a noise. Peter opened his eyes. He saw that the room had grown dark. Something was outside the window.

Tap, tap, came a noise at the window. "What is it?" called Peter. "It is I, Peter. Come to the window." Peter climbed out of bed and went to the window. There he saw a bird. This bird was so large that his eyes were as big as saucers.

"Hello, Peter," said the bird. "Hello," said Peter. "Who are you?" "I am a bird, Peter. I have come to take you to the moon. I heard your father say that birds cannot fly to the moon. But he forgot about fairy birds. I am the King of the Fairy Birds."

Tests for Comprehension

Since eye-movements do not indicate the presence or absence of comprehension, some measure is ordinarily used to indicate whether the reader has succeeded in interpreting the material, or whether he has simply allowed his eyes to skim the printed line. Several methods are available, but probably the type best suited to deaf children is that of having them respond to questions whose answers require a minimum use of language. The scores on this test are valuable only to indicate that comprehension is present. A

standardized test is made use of to measure the subjects' ability to read printed material.

The deaf child may have difficulty in understanding the questions that are intended to test comprehension of material read. He may fail to answer, not because he does not know the content of the reading material, but because he does not understand the language of the question form. The problem of formulating a comprehension test for this investigation was something more than merely formulating questions on the content of the test material. The questions for the comprehension tests used in the present study were chosen by the following procedure.

The experimenter formulated a series of fifteen questions from the content of the more advanced experimental selection and ten from the easier selection. The experimental selections were then presented to twenty-three deaf children who were not to participate in the experiment (thirteen in the first and second grades and ten in the third grade). They were not told the purpose of the procedure, but simply told to read the sheets handed them. After all the subjects had finished reading, the mimeographed lists of questions were given them, and they were told to write their responses. From these responses a tabulation was made of the number of times each question had been answered correctly. The six questions receiving the greatest number of correct responses were chosen for the final tests, on the assumption that those receiving the greatest number of correct responses were questions which fitted the comprehension ability of the subjects both for the reading material and for the form of the question.

Apparatus

The eye-movement technique has been used for a number of years at The University of Texas, and the camera used in the present investigation has been described in several experimental studies. A clear understanding of the apparatus can be obtained by reading the combined descriptions of Gray[16] and Holland.[17]

The camera was taken to the Texas School for the Deaf and installed in one of the teachers' rooms of the Main Building. Having the camera at the school was advantageous for several reasons. It gave the students an opportunity to become acquainted with the camera before their turn came for reading, and a feeling of ease which they probably would not have had if they had been required to make the trip to the University laboratory.

Presence or Absence of Growth in Three Fundamental Elements of Reading

In Table II is listed the average number of fixations per line, the average duration of each fixation, and the average number of regressive movements per line for each of the ninety-four deaf readers who took part in this experiment. In this table is also given the grade placement of each subject, as well as his total reading score on the New Stanford Reading Test and the grade norm to which the score corresponds. From these data grade medians for the measurements of eye movements were computed.

Figure 1 shows in graphic form the position of the medians for the number of fixa-

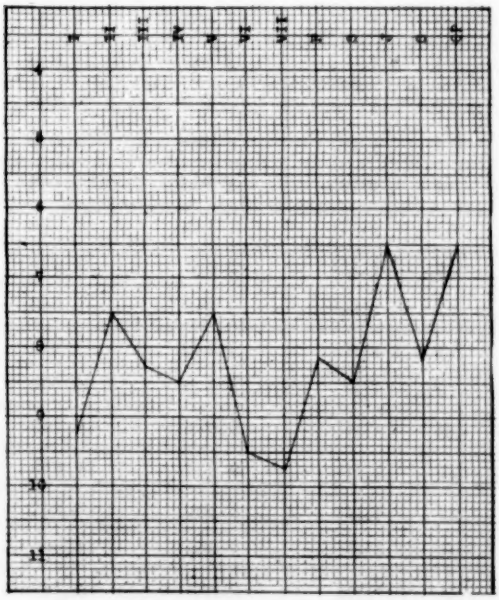


Fig. 1. Relationship between the Average Number of Fixations per Line and School Grade. The school grade is shown on the horizontal axis; the average number of fixations per line is shown on the vertical axis.

[16] C. T. Gray, *Deficiencies in Reading*, pp. 179–182. Boston: D. C. Heath and Company, 1922.
[17] Benjamin F. Holland, "Special Apparatus in the Laboratory of Educational Psychology, University of Texas," *American Journal of Psychology*, XLV (January, 1933), 139–145.

TABLE II*

DATA FOR READING OF ALL SUBJECTS

Subject	School Grade	Average Number of Fixations per Line	Average Duration of Fixations	Average Number of Regressive Movements per Line	Total Reading Score on the New Standard Achievement Test	School Grade Norm for Reading Score
1	IB	7.400	9.735	.20	22	2.6
2	IB	4.250	8.588	.00	24	2.7
3	I	6.600	9.515	.00	36	3.2
4	I	10.200	9.666	1.40	33	3.1
5	I	9.400	11.829	.40	38	3.3
6	II	7.200	10.750	.40	18	2.5
7	II	7.600	8.868	1.20	29	2.9
8	II	4.400	7.500	.40	27	2.8
9	II	7.200	9.055	.20	26	2.8
10	II	7.800	9.538	.40	35	3.2
11	II	8.800	11.840	.20	33	3.1
12	II	7.400	10.189	.80	17	2.4
13	IIIE	9.692	12.674	2.23	25	2.8
14	IIIE	9.384	9.606	2.23	35	3.2
15	IIID	7.111	6.705	1.45	32	3.1
16	IIID	7.461	6.278	1.23	31	3.0
17	IIID	9.363	9.563	.73	29	2.9
18	IIIC	7.769	7.287	.15	33	3.1
19	IIIC	8.142	17.000	.14	37	3.3
20	IIIC	9.076	10.169	.31	39	3.4
21	IIIB	9.000	11.615	.85	28	2.9
22	IIIB	6.000	8.644	.50	33	3.1
23	IIIB	9.090	11.620	1.73	37	3.3
24	IIIB	8.083	9.247	1.75	35	3.2
25	III	9.538	16.322	.46	36	3.2
26	III	7.538	11.051	.85	30	3.0
27	III	7.750	10.354	2.00	20	2.6

*In computing all averages the first and last lines of each paragraph were omitted.

TABLE II* (Continued)

DATA FOR READING OF ALL SUBJECTS

Subject	School Grade	Average Number of Fixations per Line	Average Duration of Fixations	Average Number of Regressive Movements per Line	Total Reading Score on the New Standard Achievement Test	School Grade Norm for Reading Score
28	III	5.777	8.365	.78	29	2.9
29	III	6.307	8.195	.31	31	3.0
30	4-1A	8.714	10.737	1.38	29	2.9
31	4-1A	6.250	6.800	1.11	39	3.4
32	4-1A	12.400	8.241	2.00	29	2.9
33	4-1A	9.000	10.155	.90	39	3.4
34	4-1A	10.333	9.633	2.67	35	3.2
35	4-1A	9.666	9.793	.20	31	3.0
36	4-1A	7.400	8.297	.60	31	3.0
37	4-2A	9.000	8.777	.77	40	3.4
38	4-2A	9.000	8.931	.62	35	3.2
39	4-2A	8.153	6.226	1.62	39	3.4
40	4-2A	8.250	12.848	1.00	36	3.2
41	4-2A	9.333	12.089	1.34	24	2.7
42	4-3A	9.500	12.342	.50	31	3.0
43	4-3A	8.833	12.333	.50	47	3.9
44	4-3A	7.461	11.175	.54	42	3.6
45	4-3A	7.250	8.564	.42	47	3.9
46	5-1A	7.153	8.591	2.00	28	2.9
47	5-1A	7.750	10.838	.00	32	3.1
48	5-1A	6.000	8.166	.22	47	3.9
49	5-1A	6.400	8.843	.10	48	4.0
50	5-1A	9.545	7.276	2.55	50	4.1
51	5-1A	8.153	9.641	1.62	37	3.3
52	5-1A	9.076	7.288	1.00	41	3.5
53	5-1A	7.153	7.967	.08	42	3.6
54	5A	8.230	9.168	.62	44	3.7
55	5A	8.000	8.028	1.00	44	3.7

*In computing all averages the first and last lines of each paragraph were omitted.

TABLE II* (Continued)

DATA FOR READING OF ALL SUBJECTS

Subject	School Grade	Average Number of Fixations per Line	Average Duration of Fixations	Average Number of Regressive Movements per Line	Total Reading Score on the New Standard Achievement Test	School Grade Norm for Reading Score
56	5A	7.181	9.051	.36	31	3.0
57	5A	8.923	8.603	1.15	29	2.9
58	6A	8.000	7.423	1.08	60	4.7
59	6A	9.538	8.588	.62	26	3.9
60	6A	10.800	10.672	1.20	45	3.8
61	6A	9.000	16.837	.77	46	3.9
62	6A	9.222	9.337	1.11	50	4.1
63	7-1A	6.846	8.806	.62	50	4.1
64	7-1A	10.000	11.818	1.75	45	3.8
65	7-1A	10.153	8.393	1.08	53	4.3
66	7-1A	9.400	7.127	2.30	64	5.1
67	7-1A	7.076	9.293	.38	56	4.4
68	7-1A	10.230	9.736	1.31	56	4.4
69	7-1A	9.230	9.350	.62	43	3.6
70	7-2A	7.461	10.309	.38	44	3.7
71	7-2A	9.307	9.719	.92	62	4.9
72	7-2A	10.307	12.910	.92	47	3.9
73	8-1A	9.916	14.362	.75	58	4.6
74	8-1A	8.769	9.236	.23	47	3.9
75	8-1A	8.461	8.418	.38	65	5.2
76	8-2A	7.333	9.859	.44	67	5.4
77	8-2A	7.461	10.268	.92	61	4.8
78	8-2A	8.076	6.304	.08	58	4.6
79	8-2A	6.307	7.634	.31	53	4.3
80	9-1A	8.819	8.618	1.00	63	5.0
81	9-1A	6.857	7.291	.14	60	4.7

*In computing all averages the first and last lines of each paragraph were omitted.

TABLE II (Continued)

DATA FOR READING OF ALL SUBJECTS

Subject	School Grade	Average Number of Fixations per Line	Average Duration of Fixations	Average Number of Regressive Movements per Line	Total Reading Score on the New Standard Achievement Test	School Grade Norm for Reading Score
82	9-1A	10.090	9.513	1.09	49	4.0
83	9-1A	8.916	6.598	.75	71	5.7
84	10-1A	6.250	9.373	.25	65	5.2
85	10-1A	6.846	7.629	.92	79	6.6
86	10-1A	6.923	8.224	.62	60	4.7
87	Sen.-1	6.538	7.305	.31	85	7.4
88	Sen.-1	7.000	8.642	.14	74	6.0
89	Sen.-1	8.846	8.017	1.46	64	5.1
90	Sen.-1	9.307	9.173	1.00	78	6.4
91	Sen.-1	6.909	7.638	.58	72	5.8
92	Sen.-2	8.230	9.121	.77	63	5.0
93	Sen.-2	8.538	9.675	.08	55	4.4
94	Col. Prep.	6.000	5.296	.00	85	7.4

tions per line with respect to grade levels. The degree of fluctuations from grade to grade seems to indicate no tendencies for development by grade levels for this particular element of reading, but certain trends are discernible.

The graph in Figure 2 indicates that the relationship between school grade and average duration of fixations is also a fluctuating one. Although they show a trend toward maturity in this phase of the reading process, they exhibit no levels of growth directly correlated with advancement in grade classification.

Rhythmic progress along the printed line is another element of the reading process which indicates development in the reading habit. The number of regressive movements of the reader's eyes serves as an index of the degree of maturity in this element of the reading process.

Figure 3 was constructed from the medians of each school grade for the average number of regressive movements per line. Grade fluctuations in this measure render analysis difficult, but, in rough outline, the graph appears to fall into two general divisions, with grade 6 as the crucial point. To this point there seems to be, on the whole, an increase in the number of regressions with advancement in grades. From this point on there is a fluctuating decrease in the average number of regressions until the highest grade level is reached.

In order to study more fully the trends suggested by the graphic relationships shown in Figures 1, 2, and 3, coefficients of correla-

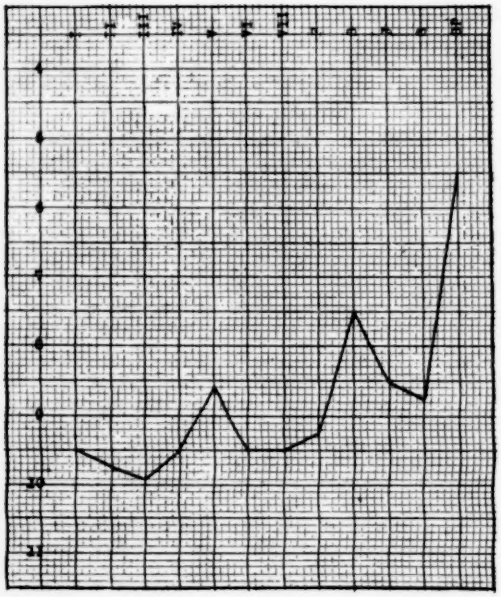


Fig. 2. Relationship between Average Duration of Fixation Pauses and School Grade Classification. The school grades are shown on the horizontal axis; the average durations of fixation pauses are shown on the vertical axis.

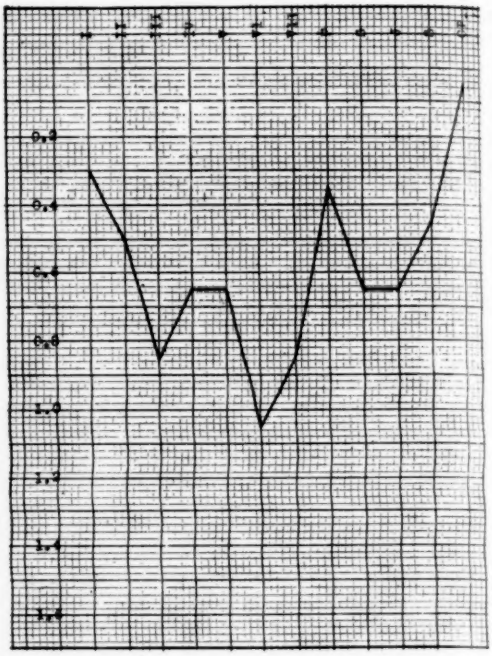


Fig. 3. Relationship between the Number of Regressive Movements per Line and School Grade. The school grade is shown on the horizontal axis; the average number of regressive movements per line is shown on the vertical axis.

tion were calculated for the three eye-movement measures and grade placement: (1) for all the subjects taken together, and (2) for two divisions of the subjects (grades 1 to 5 inclusive). These relationships are shown in Table III.

Although the coefficients of correlation are small, they may be interpreted to mean that there is a trend for the readers in the less mature group to increase both the number of fixations per line and the number of regressive movements per line until grade 5 is completed, but for the group beginning with grade 6 the tendency is to lessen both the number of fixations and regressive movements. A slight decrease in the average du-

TABLE III

RELATION OF GRADE CLASSIFICATION TO MEASURES OF EYE-MOVEMENTS

	Grades 1 to 5 Inclusive N=57	Grades 6 to Col. Prep. Inclusive N=37	All Grades N=94
Average Number of Fixations per Line	r—.143±.088	r .513±.082	r .0349±.0695
Average Duration of Fixation Pauses	r .157±.087	r .346±.098	r .183±.067
Average Number of Regressive Movements per Line	r—.180±.086	r .30±.101	r .148±.068

ration of fixations prevails for both groups of readers.

Table IV was constructed in order to compare the eye-movement habits of the deaf readers of this study with the hearing ones of the Buswell study. It will be remembered that, for the third grade level and above, the test selection used by Buswell for hearing subjects was used for the deaf readers of this study.

Table IV gives the medians for the three eye-movement habits found by Buswell for

TABLE IV

COMPARISON BETWEEN EYE-MOVEMENTS OF DEAF AND HEARING SUBJECTS

(1)

Median average for the number of fixations per line

	School Grade											
	1B	2	3	4	5	6	7	F	S	J	S	C.P.
Hearing Subjects Buswell Study	15.05	10.7	8.9	7.3	6.9	7.3	6.8	7.2	5.8	5.5	6.4	5.9
Deaf Subjects of This Study	9.25	7.5	8.25	8.50	7.50	9.50	9.67	8.17	8.50	6.50	8.17	6.50
For Deafs—Av. No. of Yrs. in School*	4	3.7	5.9	7.6	7.9	7.6	10.2	10.9	11.9	12.3	13.4	13.0

(2)

Median for average duration of fixations

	School Grade											
	1B	2	3	4	5	6	7	F	S	J	S	C. P.
Hearing Subjects	10.8	9.1	7.9	6.7	6.3	5.9	6.0	6.1	6.2	5.6	6.2	6.3
Deaf Subjects of This Study	9.5	9.75	9.87	9.50	8.60	9.50	9.50	9.25	7.50	8.5	8.75	5.5
For Deafs—Av. No. of Yrs. in School	4	3.7	5.9	7.6	7.9	7.6	10.2	10.9	11.9	12.3	13.4	13

(3)

Median for average number of regressive movements per line

	School Grade											
	1B	2	3	4	5	6	7	F	S	J	S	C.P.
Hearing Subjects Busewell Study	4.0	2.3	1.8	1.4	1.3	1.6	1.5	1.0	0.7	0.7	0.7	.50
Deaf Subjects of This Study	.30	.60	.85	.65	.65	1.05	.85	.37	.65	.65	.45	.05
For Deafs—Av. No. of Yrs. in School	4	3.7	5.9	7.6	7.9	7.6	10.2	10.9	11.9	12.3	13.4	13.0

*Buswell assumed a grade for each year in school.

the different grades of hearing subjects, and those found by the writer for the deaf subjects of this study. The deaf child is usually in school two or three years in preparatory work before he can be placed in the regular first grade; hence, the average number of years in school for each grade of the deaf readers is shown on the last row of each part of the table.

Assuming, on the average, an advancement of one grade per year for the hearing subjects, the fourth grade subjects have a school experience of four years. The fourth grade hearing readers have a median average of 7.3 fixations per line, which is approximated by deaf readers who have been in school 3.7 years; but a median average of 6.9 fixations for hearing subjects with five years of school experience is approximated by deaf readers only after 12.3 years of school experience. The average number of fixations per line for the three deaf readers in the junior class of high school, and also for the one deaf reader in the college preparatory class, approximates the average for the hearing subjects in the senior year in high school. From these data it seems that deaf readers may approach the maturity of hearing readers in this element of the reading habit, but the approach to maturity does not progress in stages related to grade levels.

The second part of Table IV shows the grade medians of hearing and deaf subjects for the average duration of fixations. The performance of the one deaf subject in the college preparatory class shows that it is possible for a deaf reader to reach the same degree of maturity in this element of the reading habit as hearing readers. The grade medians for the several grade levels, however, do not indicate growth in this element in stages comparable with school grade advancement.

The third part of the table shows the grade medians of hearing and deaf subjects for the average number of regressive movements per line for the different grades. An examination of this table reveals striking differences between the hearing and deaf readers for this element in the reading process, both as to the number of regressions and as to their relationship to grade classification. For the deaf readers there seem to be no growth stages comparable with school grade advancement.

Reading Scores

Since many factors operate in the classification of deaf students into grades, it seems desirable to correlate school grade placement with scores on a standardized reading test. The New Stanford Reading Test, Form W, was used for the first three grades, and the Reading Test of the New Stanford Achievement Test, Form X, was used for the other grades.

Figure 4 shows this relationship graphically when the score averages are plotted with school grades. The graph indicates a close

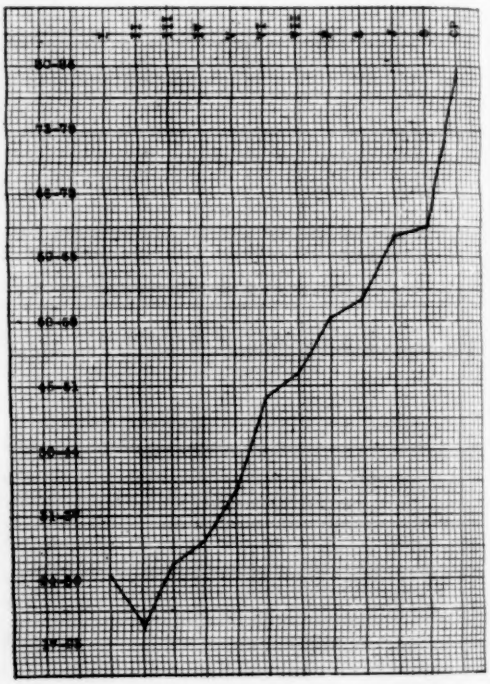


Fig. 4. Relationship between Grade Average Scores on the New Stanford Reading Test and School Grade. The intervals into which the grade averages fall are given on the vertical axis, and the school grades along with their mean scores on the reading test are given on the horizontal axis.

relationship between advancement in grade placement and ability to score on the reading test, the only instance for which this is not the case being in grade 2. The coefficient of correlation shows this relationship to be $r. = .886 \pm .015$. These data show that deaf children do increase their ability to com-

prehend printed material with advancement in school grades.

Since the data show that the eye-movements of deaf subjects do not approach those of mature readers in stages of growth comparable with school grade, but that comprehension of reading as measured by a standardized test correlates fairly high with grade classification, the question arises as to the extent to which eye-movements are related to scores in the reading test.

Figure 5 shows the relationship between the medians of the average number of fixations per line and test scores on the reading test. This graph indicates two trends of relationships with very few fluctuations. Until the readers are able to merit a reading score falling within the interval 45–51, comprehension is accomplished by an increasing number of fixations. Beyond that level of ability, deaf readers tend to decrease the number of fixations per line as the ability to interpret printed language increases.

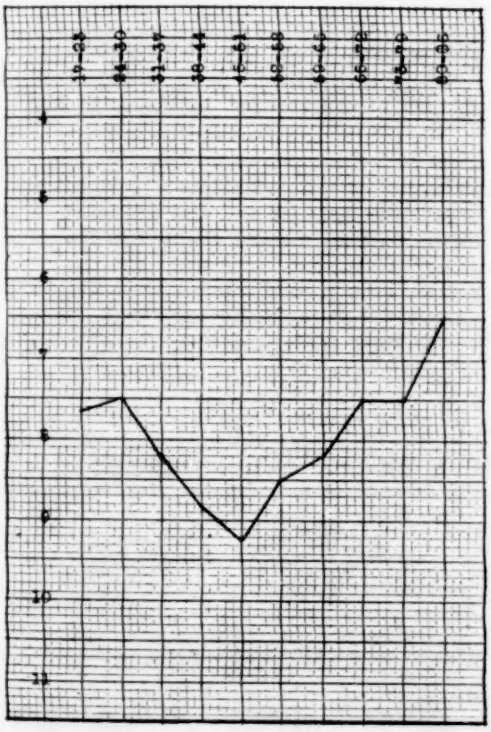


Fig. 5. Relationship between the Average Number of Fixations per Line and the Total Reading Score on the New Stanford Achievement Test. Reading scores are shown on the horizontal axis, and the average number of fixations are shown on the vertical axis.

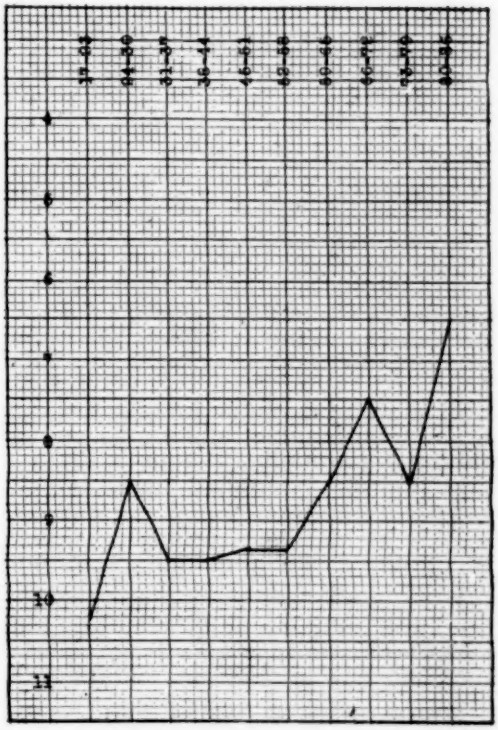


Fig. 6. Relationship between Average Duration of Fixation Pauses and Total Reading Score on the New Stanford Achievement Test. Reading score is shown on horizontal axis, and average duration of fixation pauses is shown on vertical axis.

Figure 6 exhibits graphically the relationship between the average duration of fixations and the scores on the New Stanford Reading Test, and indicates that after a score falling within the interval 31–37 is reached, there is almost a plateau in the average duration of fixations accompanying the increasing scores on the reading test through three step-intervals. After the interval 52–58 is reached, apparently increased comprehension scores are accompanied by a decrease in duration of fixation with a fluctuation on the 73–79 interval and a final rise for the subject on the highest step-interval.

Figure 7 indicates the relationship between scores made on the New Stanford Reading Test and the average number of regressive

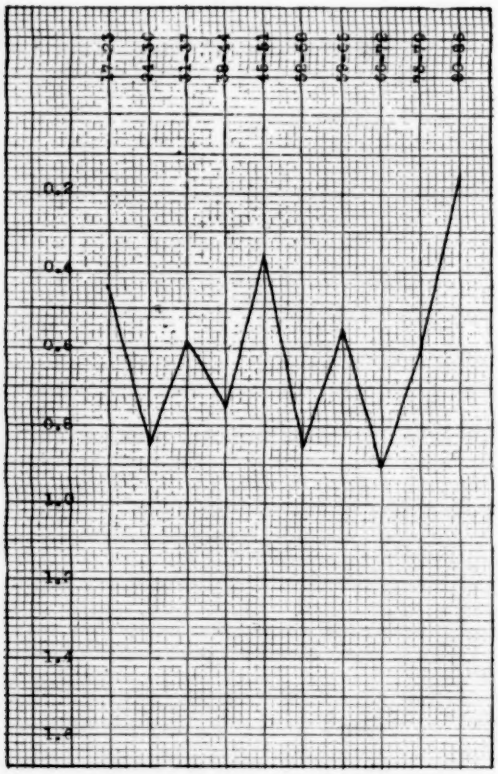


Fig. 7. Relationship between the Average Number of Regressive Movements per Line and the Total Reading Score on the New Stanford Achievement Test. The reading score is shown on the horizontal axis, and the regressive movements are shown on the vertical axis.

movements per line. The graph reveals no grade trends in relationship between the average number of regressive movements per line and the higher comprehension scores on the reading test; or it may be said that for this measure individual variations submerge any trend for comparable stages of growth in the two measures.

In order to ascertain trends which these graphs fail to indicate, coefficients of correlation were calculated for the relationship between the three eye-movement measures and the reading scores for all ninety-four subjects, and also for the two groups, grades 1 to 5 inclusive, and grades 6 to college preparatory inclusive. These coefficients are given in Table V.

When all the grades are considered, the coefficients indicate little relationship, but the coefficients, though low, for the two groupings, show trends for an increase in comprehension scores to be accompanied by an increase in the number of fixations for the first five grades, while the increase in reading score is accompanied by a decrease in the number of fixations for the higher grades.

The relationship between the average durations of fixation pauses and reading scores, as indicated by the coefficients of correlation, is practically zero for the first five grades. It seems that no marked changes in the durations of fixations are made by the lower grades, even though interpretative ability increases. This same observation may be made of the lower grades regarding the relationship

Table V

Relation of Eye-Movement Measures to Reading Scores for Two Groups of Readers

	Grades 1 to 5 Inclusive N=57	Grades 6 to Col. Prep. Inclusive N=37	All Grades N=94
Average Number of Fixations per Line	$r = .126 \pm .088$	$r = -.316 \pm .100$	$r = -.0603 \pm .069$
Average Duration of Fixations	$r = -.071 \pm .089$	$r = -.515 \pm .082$	$r = -.2903 \pm .069$
Average Number of Regressive Movements per Line	$r = -.031 \pm .089$	$r = -.231 \pm .105$	$r = -.226 \pm .066$

between regressive movements and reading scores made on the New Stanford Reading Test. For the higher grades, however, there seems to be a tendency for maturity in all three of the eye-movement measures to be accompanied by an increased score in reading ability.

Types of Regressive Movements of Various Grade Levels

A study of the records of the deaf readers indicates that the regressions present may be classified into different types. This classification has been attempted in order to determine, if possible, the part they play in the deaf subject's reading habits. It is true that this type of analysis is more or less subjective and liable to error. The types of regressions seem to fall under six main headings:

(1) *Lack of precision in backward swing.* This type occurs usually near the beginning of the line, and is of short duration.

(2) *Too long a span.* The deaf reader does not appear to make this type of regression in an attempt to make few fixations, as is the case of the mature reader who overreaches his maximum span and finds it necessary to make a backward eye-movement for clarity. The deaf reader seems to allow his eye to jump beyond the possibility of his maximum span for no apparent reason.

(3) *Lack of understanding of concept.* Such expressions as *all at once, outside, climbed out of bed, so large that, grown dark, fly to the moon,* etc., are expressions which seem to be especially confusing to deaf readers.

(4) *Grammatical reasons.* This type of regression is different from the type designated as a lack of understanding of the concept.

(5) *Compound or unfamiliar words.* It appears to be a habit with some of the readers to fixate on the last syllable of such words as *something, everything, window, forget,* etc., and then to regress to the first syllable.

(6) *Sentence form and punctuation.* This type of regression is particularly common for the sentences, *Come to the window, There he saw a bird, But he forgot about fairy birds,* and the question forms *"Who are you?"* and *"What is it?" called Peter.*

Table VI

Types of Regressive Movements for Different Grades Expressed in Percentages

Grade Mean for Comprehension Score	Grade	1 Lack of Precision in Backward Swing	2 Too Long Span	3 Lack of Understanding of Concept	4 Grammatical Construction	5 Compound and Unfamiliar Words	6 Form of Sentence and Punctuation
51.2	I	50	0	30	20	0	0
69	II	71	0	17.4	11.6	0	0
60.5	III	46.5	1	21.5	17.5	4	9.5
65.6	IV	53.6	3.2	19.2	15.2	4.8	4
69.3	V	29.7	3.1	27.3	26.6	3.1	10.2
84.6	VI	46.3	0	29.6	13	3.7	7.4
79.6	VII	46.3	1.6	25.2	16.3	2.4	8.2
75.3	VIII	31.6	2.6	21.1	31.6	0	13.1
76	IX	30.3	0	42.4	18.2	3.0	6.1
74	X	56.5	0	4.4	26.1	4.35	8.65
84.1	XI	43.4	0	35.8	17	0	3.8

An examination of Table VI shows that the highest percentage of regressions for all the grades is lack of precision in the backward swing. With the exception of grades VIII and X, a lack of understanding of meanings of phrases caused more regressions than did grammatical construction.

That the percentages of regression for compound and unfamiliar words and for sentence form and punctuation are lowest for all the grades may be due to the fact that opportunities for these types were limited. There are no grade trends for any particular type.

Summary

From the data of the study the writer is able to draw some conclusions with respect to certain questions raised at the outset of the study:

1. The deaf subjects do not show stages of growth in eye-movements parallel with advancement in grade classification.

2. Deaf readers show maturity levels in eye-movements, but these levels are not attained "essentially along the same lines" as those in the eye-movements of hearing readers.

3. Coefficients of correlation between grade classification and eye-movement measures indicate a tendency for readers in the first five grades to increase both the number of fixations and the number of regressive movements, but for readers in the higher grades to decrease both the number of fixations and the number of regressive movements.

4. A slight decrease in the average duration of fixations prevails for both groups of readers.

5. The deaf readers attain increased interpretative ability with advancement in grade classification but in a retarded way. The highest reading score made by a deaf reader in the college preparatory class corresponds to a reading grade of 7.4.

6. The coefficients of correlation between eye-movement measures and reading scores on a standardized test show a tendency for an increase in comprehension to be accompanied by an increase in the number of fixations for the first five grades, while for the higher grades the increase in the reading score is accompanied by a decrease in the number of fixations.

7. The coefficients of correlation indicate that, for the higher grades, maturity in the three phases of eye-movements considered accompanies growth in interpretative ability.

8. Deaf readers show a coefficient of correlation between the average number of fixations and the average duration of fixations of $.372 \pm .0599$. This is slightly lower than the coefficient of correlation between these measures for hearing subjects in grades 2 to 6 inclusive, but higher than the coefficient of correlation between these measures for hearing subjects at the high school junior class level.

9. A comparison of the coefficients of correlation between the number of fixations and the duration of fixations for grades 1 to 5 inclusive, and for grades 6 to 11 inclusive, shows practically no difference in the relationships between these two eye-movement measures for the two groups.

10. Deaf readers make fewer regressive movements than hearing readers. The regressions made by deaf readers appear to fall into different types which are produced by as many different causes. The most common type of regression for all the grades is that which occurs at the beginning of a line when the return sweep of the eye fails to carry the fixation back to a point at which the initial part of the line is clear.

11. In all grades except 8 and 10, the failure to understand concepts expressed by language symbols accounts for a higher percentage of regressions than does the need for grammatical analysis.

12. All the grades except grade 4 show a higher percentage of regressions due to sentence form and punctuation than is due to compound and unfamiliar words.

13. There are no grade trends for any particular type of regressive movements.

References

Allport, Floyd Henry. *Social Psychology*, pp. 169–98. Boston: Houghton Mifflin Company, 1924.

Book, William F. "The Development of Higher Orders of Perceptual Habits in Reading," *Journal of Educational Research*, XXI (March, 1930), 161–76.

Brown, Andrew W. "The Correlation of Non-Language Tests with Each Other, with School Achievement, and with Teachers' Judgments of the Intelligence of Children in a School for the Deaf", *Journal of Applied Psychology*, XIV (1930), 371–75.

Buell, Edith. *Outline of Language for Deaf Children with Notes on Presentation of the Work.* New York: B. D. Benson and Sons, Inc.

Buswell, Guy T. *A Laboratory Study of the Reading of Modern Foreign Languages.* Publications of the American and Canadian Committee on Modern Language. New York: The Macmillan Company, 1927.

Buswell, Guy T. *Fundamental Reading Habits: A Study of Their Development.* Supplementary Educational Monographs, No. 21. Chicago: University of Chicago, 1922.

Gates, Arthur I. "A Study of the Role of Visual Perception, Intelligence, and Certain Associative Processes in Reading and Spelling," *Journal of Educational Psychology,* XVII (1926), 433–45.

Gates, Arthur I. "Methods and Theories of Learning to Spell Tested by Studies of Deaf Children", *Journal of Educational Psychology,* XVII (1926), 289–300.

Gray, C. T. *Types of Reading Ability As Exhibited Through Tests and Laboratory Experiments.* Supplementary Educational Monographs, No. 5. Chicago: University of Chicago, 1917.

Gray, C. T. *Deficiencies in Reading.* Boston: D. C. Heath and Company, 1922.

Gray, W. S. *Remedial Cases in Reading: Their Diagnosis and Treatment.* Supplementary Educational Monographs, No. 22. Chicago: University of Chicago, 1922.

Holland, Benjamin F. "Special Apparatus in the Laboratory of Educational Psychology, University of Texas", *American Journal of Psychology,* XLV (1933), 139–45.

Judd, Charles H., and Buswell, Guy T. *Silent Reading: A Study of the Various Types.* Supplementary Educational Monographs, No. 23. Chicago: University of Chicago, 1922.

Judd, Charles H. *Reading: Its Nature and Development,* pp. 54–81. Supplementary Educational Monographs, No. 10. Chicago: University of Chicago, 1918.

Knight, James. "A Laboratory Study of the Reading Habits of Spanish-Speaking Children". Unpublished Doctor of Philosophy Dissertation, The University of Texas, 1931.

Long, John A. *Motor Abilities of Deaf Children.* Teachers College Contributions to Education, No. 514. New York: Teachers College, Columbia University, 1932.

MacKane, Keith. *A Comparison of the Intelligence of Deaf and Hearing Children.* Teachers College Contributions to Education, No. 254. New York: Teachers College, Columbia University, 1927.

O'Brien, John Anthony. *Reading: Its Psychology and Pedagogy.* New York: The Century Company, 1926.

Pintner, R., and Paterson, D. G. "Psychological Tests of Deaf Children", *Volta Review,* XIX (December, 1917), 661–67.

Pintner, R., and Paterson, D. G. "The Binet Scale and the Deaf Child", *Journal of Educational Psychology,* VI (April, 1915), 201–10.

Pintner, R., and Paterson, D. G. "A Class Test with Deaf Children", *Journal of Educational Psychology,* VI (1915), 591–600.

Pintner, R., and Paterson, D. G. "Learning Tests with Deaf Children", *Psychological Monographs,* Vol. XX, No. 4, January, 1916.

Pintner, R., and Paterson, D. G. "A Measurement of the Language Ability of Deaf Children", *Psychological Review,* XXIII (November, 1916), 413–36.

Reamer, Jeanette Chase. "Mental and Educational Measurements of the Deaf", *Psychological Monographs,* Vol. XXIX, No. 3, 1921.

Starch, Daniel. *Educational Psychology.* New York: The Macmillan Company, 1928.

Thompson, Helen. *An Experimental Study of the Beginning Reading of Deaf Mutes.* Teachers College Contributions to Education, No. 254. New York: Teachers College, Columbia University, 1927.

Watson, John B. *Behaviorism,* Chapter X. New York: W. W. Norton and Company, Inc., 1925.

Weiss, Albert Paul. *A Theoretical Basis of Human Behavior,* pp. 307–27. Columbus, Ohio: Adams and Company, 1929.

"Do Deaf Children Grasp Abstractions Readily?" *Volta Review,* XXXIII (1931), 75–78. An Abstract of the experiment reported by Dr. Hofler, *Zeitschrift für Paedagogische Psychologie,* 1927.

"Experiments on the Thinking of the Deaf," *Volta Review,* XXXIII (1931), 313–15. A resumé of an article by Wilhelm Frohn in *Archiv für die Gesamte Psychologie.*

VISUAL DEFECTS AS FACTORS INFLUENCING ACHIEVEMENT IN READING

L. P. Farris
Principal, Oakland High School, California

The purpose of this study was to determine whether visual defects influence the achievement in reading of school children of the seventh grade in the Oakland Public Schools. In particular, answers have been sought to the following questions: (1) Do different types of eye defects when studied separately affect achievement in reading? (2) Do eye defects when considered collectively influence achievement in reading? (3) Are defective eyes when optically corrected more or less proficient than defective eyes not provided with lenses, as evidenced through achievement in reading? (4) Are defective eyes when optically corrected generally more or less proficient than normal eyes, as evidenced through achievement in reading?

The procedure in the collection of data involved the giving of intelligence tests, the administration of tests for visual efficiency, and the giving of achievement tests in reading at the beginning and end of the study. The Kuhlman–Anderson Intelligence Test and the Stanford Achievement Test in Reading were given late in October, 1930, to approximately 1900 H-6 grade pupils. From these tests, summaries of which were on file at the Department of Research, Oakland Public Schools, the chronological ages, intelligence quotients, and reading scores were obtained for use later in the study. In January and February, 1931, under the immediate direction of the Division of Optometry, University of California, visual efficiency tests including tests for the efficiency of the functions of accommodation and convergence were given to 1685 pupils of the L-7 grade in the 16 public junior high schools of Oakland.

For many years school procedures have been based upon the assumption that vision is free from impediments if, and when, in perceiving distant objects, the acuity equals or approximates the value generally accepted as normal. The adoption of a standard on the assumption that the rating in visual acuity is a measure of the ease with which the eyes perform their visual functions is to accept a standard which places major emphasis upon retinal activity and to a great extent ignores the muscular functions of the eye which are known to lie at the basis of the numerous discomforts which frequently affect the eyes in vision. Accommodation of the eyes, and the convergence of their visual axes are functions essential to the acts of vision and require in most phases of their activity definite contractions of muscle groups. The visual acts of reading are impossible without the coordination of these muscularly controlled functions, and the conditions of comfort which attend the reading process definitely depend upon the degree of perfection with which they act. For this reason, in this study, tests were given and data obtained indicating the efficiency of the functions of accommodation and convergence in addition to those tests which primarily disclose the nature of the focusing of the eyes.

The study is delimited to a consideration of visual efficiency as it affects achievement in reading for a period of one year by the pupils in the seventh grade in the Oakland Public Schools—to be exact, 14 months, October, 1930, to December, 1931. The procedure was further delimited and refined by selecting pupils for the control group whose chronological age and mental ability were approximately equivalent to those of the defective group.

The method was that of comparing statistically the achievement in reading of seventh grade pupils, equal in ability and chronological age, but different in visual acuity. The study was conducted under reasonably well controlled conditions; this statement is borne out by the fact that the difference in average chronological age between the normal and defective groups never exceeds six months, usually one or two months, and the difference in average intelligence quotients is more than

two in but one instance; moreover the average chronological age for the entire 384 pupils of the control group was identical to that of the 384 in the defective group; also, there was only 0.2 of one unit difference in average I.Q.'s of the groups. Furthermore, any inequalities which might be expected from the teacher factor were presumably ironed out since there were 95 different teachers teaching the 768 pupils, 46 while the pupils were in the L-7 grade and 49 different teachers in the H-7 grade. For instance, there was an average of less than three pupils per teacher for each type of eye defect involved in the study.

The major findings of this study are:

Forty-four per cent, i.e. 739 of the 1685 pupils in this study who were given the visual tests, were found to have eye defects of varying degrees. For purposes of comparison the White House Conference Reports on Child Health and Protection estimate that approximately 19 per cent of the school population have eye defects which can be remedied so that the children so affected may be brought within the normal group for educational purposes; the Oakdale Survey shows that 63% were affected with some degree of refractive error; Berkowitz reports that estimates approximate 42% of school children as having eye defects; Hilleboe's study gives 12% as serious enough to be classed as health defects.

Both hyperopia and strabismus are associated with less than normal progress in reading; while myopia and myopic astigmatism were both found to be associated with more than normal progress. The hypothesis that greater strain on the nervous system is required to adjust hyperopic eyes to the conditions which must be satisfied in reading seems confirmed by the finding of this study that pupils with normal eyes have 97 chances in 100 to make greater gains in reading than do pupils affected with varying degrees of hyperopia (critical ratio + 2.67, 78 cases studied.) Also the less generally accepted hypothesis that myopic eyes adjust themselves to reading with less exertion of the muscles of accommodation and less expenditure of nerve energy than do emmetropic eyes was rendered more tenable by this study. It was found that for the 136 pupils affected with myopia and myopic astigmatism there are 96 chances in 100 that the true difference in favor of the defectives is greater than zero (critical ratio −2.57.)

Pupils whose visual perception is monocular make progress in reading superior to those not having correct coordination of the two eyes.

Types of eye defects other than the myopic, hyperopic, and the strabismic types have little effect upon progress in reading. In other words, the data of this study do not bear out the hypothesis that children with defects in visual acuity regardless of type are always handicapped in the learning of reading. The fact that achievement in reading is often affected by muscular eye functions is believed to account for the progress in reading achievement frequently found among those pupils of this study with a low visual acuity rating, and argues in favor of the adoption of a more comprehensive standard than the visual acuity rating in appraising the efficiency of the visual functions.

When all types of eye defects were considered collectively the defectives made slightly greater gains than the normals (critical ratio −1.10). It is possible, however, that extreme cases among the one hundred thirty-six who were affected with myopia and myopic astigmatism are largely accountable for the results.

Correction lenses, in cases involving types of eye defects found to be associated with poor reading (both hyperopia and strabismus) are aids to achievement. For example, hyperopes wearing correction lenses made an average individual gain in reading score of 7.0 while an equal number of hyperopes not wearing lenses made an average of individual gains of only 5.5; for strabismus 11.2 for those with lenses against 7.3 for those without lenses. Those groups of pupils affected with myopia, myopic astigmatism, monocular visual perception, and ocular muscle imbalance, despite their lack of lenses, made a greater average of individual gains in reading score than did those defectives having the same refractive errors but wearing lenses. The myopes, for example, who were wearing lenses made an average individual gain in reading score of 7.3 while those not wearing lenses made an average gain of 9.0; for myopic astigmatism 9.1 for those with lenses against 9.6 for those without lenses. Therefore, since the myopic groups without lenses made progress superior to equivalent groups having the same errors and wearing lenses, the find-

ings indirectly confirm one of the major conclusions of this investigation; namely, that the generally accepted standard of eye structure (emmetropia) does not necessarily imply superiority of efficiency in reading. In other words, the same treatment that causes eyes to function more nearly as emmetropic eyes, such as the placing of lenses on pupils affected with myopia, was not found to improve significantly their reading efficiency.

Certain important educational implications may be drawn from this study:

The superficial eye tests (with the Snellen Charts alone) which are at present being given in many of our public schools are inadequate and should be replaced by more thorough and more complete visual tests.

The establishment and maintenance of eye clinics as a part of the health program in the public schools would enable many pupils to take better advantage of the educational program. In these clinics the aim should be to determine conditions which indicate unfavorable functioning of the eyes in so far as they pertain to school work, and to institute measures to remove such impediments to school progress. In some cases it might mean the constant use of lenses to correct refractive errors; in other cases it might necessitate lenses which have the effect of better coordinating the visual functions of accommodation and convergence, while in still other cases it might merely involve exercises for the purpose of stimulating the desired response of the eyes in vision.

There is a need for closer coordination among school counselors, school nurses, and school doctors. School counselors should be required to secure specific knowledge regarding the character of a student's visual conditions, since such information is as much a working tool for better educational and vocational guidance as is a knowledge of his I.Q., chronological age, reading age, and such other factors as are at present used by school counselors in their guidance programs.

Class room methods and procedures in teaching reading should be modified to meet the special needs of those pupils who have eye defects. Hyperopes might well be seated in the rear of the class room, whereas, if myopes must be in the same class room, they should be in the seats quite near the blackboard. The teacher should make certain that script writing on the front blackboard is large enough to be read easily by pupils in the back rows. The findings further suggest to those engaged in the administration of our schools the need of providing more suitable reading materials for those whose visual acuity is low; and, also, for those having defective eye structure even though their acuity of vision in reading may be equal or superior to those having emmetropic eyes. Books in large type, typewriters with large type, heavy lead pencils, maps without detail and other instructional material that may be used without causing eye fatigue should be provided as a part of a school system's remedial program.

A TECHNIQUE FOR DETERMINING THE RELATIVE DIFFICULTY OF WORD MEANINGS AMONG ELEMENTARY SCHOOL CHILDREN

HARVEY C. TILLEY
Supervising Principal, Franklin Borough Public Schools, Conemaugh, Pa.

I. INTRODUCTION

In order to determine the relative difficulty of a list of word concepts, it becomes necessary to know whether or not children experience difficulty in understanding the idea conveyed by the word symbol. The importance of this fact is shown by the desire to study word meanings on the part of many outstanding educational experimentalists. A very concise expression of what has been done in this field is given by Dolch[1]. Kelley[2] found that the multiple-choice type of test is the most valid and reliable instrument at present by which to determine the knowledge of words. However, the construction of a multiple-choice test in vocabulary which would involve from ten to twenty thousand words, or even one thousand words, is too tremendous an undertaking to be practical. This makes the subject of the relative difficulty of word meanings an impractical one unless a different technique can be used for testing word difficulty.

1. *The Problem*

The study reported in this paper[3] was an attempt to investigate the possibility of using a self-appraisal test as a tool for determining the relative difficulty of word meanings known by children in a given grade or group of grades in the elementary school, not only when those words occurred in isolation but also when they occurred in context.

2. *The Data*

The experiment was conducted during the second semester of the school year from January, 1934 to June, 1934 with words found in the Survey Tests of Vocabulary[4]; the Lower Extension of the Inglis Test of Vocabulary[5]; the first eighty words of the vocabulary section of the Stanford Achievement Test[6]; and the words selected as difficult from readers of the third, fifth, and seventh grades in school.

The words were converted into two types of tests, the regular multiple-choice type as found in the standardized forms and the self-appraisal type. They were then presented as tests to 326 children of the third grade, 1096 children of the fifth grade, and 1434 children of the seventh grade. Each child took both tests, the self-appraisal first, on each series of words. The words were presented to the children in both techniques at the same sitting.

The word which a child did not know was considered difficult for that child and the number of children who did not know a word constituted the difficulty score for the word. The difficulty scores received for each word when administered by the two techniques served as the data for this study.

3. *Technical Treatment of Data*

Since context often suggests the meaning of a word, each word presented to the children

---

[1] Edward W. Dolch, *Reading and Word Meanings*, Chap. I, pp. 1–23. Boston: Ginn and Company, 1927.

[2] V. H. Kelley, *An Experimental Study of Certain Techniques for Testing Word Meanings*. Doctor's Thesis, University of Iowa, 1931.

[3] The paper summarizes a thesis submitted in partial fulfillment of the requirements for the Doctor's degree at the University of Pittsburgh. The writer wishes to express his obligation to Dr. G. A. Yoakam for his assistance in this investigation.

[4] Survey Tests of Vocabulary, Forms X and Z, Grades 3–13; prepared for the Nation Wide Study of Vocabulary by L. J. O'Rourke, Director of the Study; The Survey was sponsored by the Psychological Corporation of Washington, D. C.

[5] Lower Extension of the Inglis Tests of English Vocabulary, Form Y, Grades 6–12; prepared by C. M. Downing of the Harvard Graduate School of Education; published by Ginn and Company.

[6] Stanford Achievement Test, Advanced Examination, Form A, Vocabulary Section; prepared by Kelley, Ruch, and Terman; published by the World Book Company.

tested in this study is found in a phrase, a sentence, or a paragraph.

The difficulty scores of the words used in this investigation were treated by the use of tabular and statistical analysis. Almack[7] states that the validity of a measuring instrument can be determined by checking it against other measuring instruments of known validity.

In order to secure numerical evidence of the validity of the self-appraisal technique, the multiple-choice technique was used as the known criterion and the coefficient of correlation between the techniques was found.

A series of linear correlations was made and coefficients were found for all results obtained from the two techniques of presentation, holding constant at different times the age, grade, intelligence quotient, sex, nouns, verbs, adjectives, words peculiar to the school subjects of history and science, and words thought to be difficult in ordinary reading textbooks.

II. Results of the Investigation

1. *Results with Varying Groups of Children*

In an attempt to consider the self-appraisal technique from the viewpoint of validity with pupils within a given school grade, the third, fifth, and seventh grades of various school systems were used. The several groups of words were given to these pupils in both techniques. Only the words of the Stanford-A test were given to the third grade but the coefficient of correlation was .922 ± .011. In the fifth grade all four groups of words were given, and in three cases the coefficients were above .89. In the seventh grade three groups of words were given and in all of them the coefficients were .74 or better.

The data were manipulated so that the difficulty scores of each word could be classified as the results from pupils of nine, eleven, and thirteen years of age. When the difficulty scores of the two techniques were correlated, in only one case did the coefficients fall below .74. In that case (one of age 11) it was .511 ± .049.

The pupils were divided according to their sex and correlations were made for each group of words. There was no variation in the coefficients of the sexes and all coefficients were .864 or better except with one group of words. In this group the coefficients were .583 for the males and .594 for the females.

The effect of intelligence among pupils upon self-appraisal was calculated on the basis of scores from three intelligence groups. All children with an I.Q. of 115 and above were considered as superior; all children with an I.Q. between 95–105 were considered as normal; and all children with an I.Q. of 85 and below were considered as subnormal. Three groups of words were used with these pupils. The following table reveals the results.

Table I

Correlation Between the Self-Appraisal and Multiple-Choice Techniques When the Scores are Grouped According to Intelligence

Group	Pupils	r	P. E.
Survey X, Test			
Superior	48	.991	±.001
Normal	59	.874	±.015
Subnormal	54	.863	±.017
Survey Z, Test			
Superior	40	.974	±.003
Normal	56	.951	±.007
Subnormal	34	.882	±.015
Inglis Y, Test			
Superior	43	.525	±.048
Normal	70	.358	±.058
Subnormal	40	.225	±.063

2. *Results with Varying Groups of Words*

The effect of the different types of word lists used in the experiment upon self appraisal was considered. The Survey tests were constructed for the purpose of determining a phase of English usage that might need more attention in the program of the school. The words in these tests were a sampling of a great number of words in common use. The correlation coefficients for the techniques were .877 ± .015 for Form X, and .936 ± .008 for Form Z. The words used in the

[7] J. C. Almack, *Research and Thesis Writing*, p. 67. Boston: Houghton Mifflin Company, 1930.

Stanford test were gathered for the purpose of determining the size of vocabulary for younger children. The coefficient for these words was .879 ± .017. The Inglis test was constructed for the purpose of measuring the pupil's knowledge of words not found in the everyday vocabulary, or in technical vocabularies, but often used by the general reader. When these words were used the coefficient of correlation was .661 ± .037.

Three lists of words were prepared and administered in order to determine the effect of different parts of speech upon self-appraisal. The coefficients of correlation were as follows: for nouns, .76; for verbs, .97; and for adjectives, .95. The effect of using technical words in the elementary subjects was studied by using words ordinarily found only in history and words ordinarily found only in science. The results were correlated and it was found that the coefficients were: for history, .83; and for science, .85.

3. *Results with Words Taken from Context*

A number of children in each of the third, fifth, and seventh grades of the elementary school were selected and requested to read carefully a specified story from a reader listed for their grade and not previously read by them. These children were directed first to read a paragraph and then to write the words which they did not understand upon a sheet of paper.

The words were then tabulated by the writer and their frequency of occurrence was determined. The frequency of difficulty was recorded as the difficulty score for that word. From 300 words listed by each grade, 50 words were selected and converted into a multiple-choice test. As little time as possible was permitted to intervene between the collection of the words and the presentation of the tests to the children. The difficulty scores for each technique were then correlated and the results were as follows: third grade, .534 ± .067; fifth grade, .775 ± .034; and seventh grade, .825 ± .03.

III. Findings and Conclusions

1. *Findings*

a. The validity of the self-appraisal technique was affected very little by the school grade of children but there is a tendency for a higher validity in the upper grades of the elementary school.

b. There was a very slight increase in the coefficient of correlation when a larger number of scores were made available for each word.

c. The validity of self-appraisal was not affected to any great extent when administered to children at various age levels.

d. The data suggest as a possible reason to explain why the younger children produced exceptionally high results that many of the words tested were so difficult for them as to involve no difficulty in recognizing that they did not know the meanings.

e. The validity of self-appraisal was not affected by the sex of the pupils tested.

f. The validity of self-appraisal was affected by the general intelligence of the children. The higher the intelligence of the group, the more valid the self-appraisal technique became.

g. There appeared to be a slight difference in the validity of self-appraisal depending upon whether the words tested are gathered for their value in ordinary usage or for their portrayal of concepts in reading.

h. While it did affect the coefficient of correlation to some extent, the part of speech of a word (whether noun, verb, or adjective) did not invalidate the self-appraisal technique.

i. There was no difference in the validity of self-appraisal as a result of testing words used ordinarily in history or words used ordinarily in science.

j. Even when children are requested merely to write the difficult words from the printed page, self-appraisal appeared to be valid enough to indicate word concept difficulty for children in the elementary school.

k. The word lists analyzed in the study indicated that correlations of .65 or better are worthy of consideration when questioning the validity of self-appraisal.

The facts enumerated above are conveniently summarized in Table II.

2. *Conclusions*

Thirteen correlations were summarized by Pintner[8] showing the relationship between tests which measure abstract intelligence. All of these intelligence tests are considered valid ones. They include tests constructed by

[8] Rudolph Pintner, *Intelligence Testing: Methods and Results*, pp. 56–62. New York: Henry Holt and Company, 1931.

TABLE II

CORRELATION BETWEEN THE SELF-APPRAISAL AND MULTIPLE-CHOICE TECHNIQUES WHEN THE SCORES ARE GROUPED ACCORDING TO GRAND TOTALS

List of words	No. of words	Pupils	r	P. E.
Stanford-A	80	285	.879	±.017
Survey—X	100	233	.877	±.015
Survey—Z	100	238	.936	±.008
Inglis—Y	100	333	.661	±.031
Nouns	50	190	.756	±.041
Verbs	60	193	.967	±.005
Adjectives	50	194	.945	±.010
History	35	192	.832	±.035
Science	30	188	.846	±.035
Textbook—3rd	50	163	.534	±.067
Textbook—5th	50	188	.775	±.038
Text S. A.—5th	50	188	.797	±.034
Textbook—7th	50	235	.825	±.030
Text S. A.—7th	50	235	.866	±.024

Terman, Binet, Haggerty, Herring, Pintner–Cunningham, and others. The coefficients of correlation range from .53 to .98. The average correlation is about .75.

The concrete examples given in this study indicated a validity for the self-appraisal technique, even when the coefficient of correlation was .696.

Since high validity is expressed in the two instances cited above, the findings of this study as shown by Table II point toward an exceptionally high validity for the self-appraisal technique when it is used to determine the difficulty of word meanings found in a setting of context reading materials.

By utilizing the self-appraisal technique in this manner the following possibilities may become actual realities:

a. The grading of any list of words according to the difficulty found in understanding the concepts may be partially done by submitting the list to the pupils for judgment.

b. The size of the vocabulary prevalent among various groups of elementary school children may be determined without contending with the element of chance, since the judgment of pupils as to whether or not they know word meanings is highly reliable.

c. Textbooks and other reading materials may be written utilizing word lists of known difficulty for the group intended when word difficulty has been determined by this method.

d. The nature of meaningful vocabularies found among groups of elementary school children may be determined in an economical manner.

e. Teachers may use the technique in determining difficult words which need special study in ordinary classroom work.

AN EXPERIMENTAL STUDY OF THE RELATIVE MERITS OF CERTAIN WRITTEN LETTER FORMS WITH RESPECT TO LEGIBILITY, WITH SPEED AND STABILITY AS RELATED FACTORS

HAROLD O. BORAAS
Alfred University

INTRODUCTION

The general purpose in teaching the subject of handwriting is "to develop sufficient skill to enable pupils to write easily, legibly, and rapidly."[1] No scientific tests have been reported whereby any of the existing letter forms can be rated with respect to these characteristics.

Many variant forms for each letter of the alphabet are taught in schools. Some letters have forms which differ markedly from each other. Figure 1 shows such differences in form for the letters F, G, H, and T.

Figure 1.—Differences in Forms of the Letters F, G, H, and T

The study of letter forms is urgent. Gates and Brown[2] indicate the importance of the problem as follows:

> There is need for a thorough-going experimental appraisal of the writing alphabets in order to select series of letter forms that can be learned most easily and that will result both in more rapid writing and greater legibility than any of the current types. Perhaps by selecting the best elements from cursives and print-scripts, a writing alphabet may be discovered which, by combining the merits of the various existing rivals, will be superior to any one and make the learning of the two alphabets unnecessary.

An appraisal of letter forms is important for a number of reasons. They are:

1. *General Economy.*—The public has a considerable investment in the teaching of handwriting. This might be illustrated by a conversation, something like the following, between a taxpayer and a teacher: Taxpayer—"What forms of letters are taught our children?" Teacher—"The letter forms of the X system." Taxpayer—"Why do you teach these forms?" Teacher—"We have always taught them. They look the best, and they are smooth-flowing." Taxpayer—"If these letter forms have not been tested for speed and legibility, then why should we pay money for teaching them?" Teacher—"We don't know."

2. *The Many Different Systems of Handwriting Extant.*—The Palmer, Dearborn, MacLean, Bennett, and Mills systems of handwriting contain typical variant forms for the letters of the alphabet. Some authors claim that the letter forms of their systems are found to be most suitable and useful for a rapid hand. Other authors claim that their letter forms make possible efficient handwriting movements. Whether these statements are true or not is a matter of conjecture based on individual experience with such forms. Contradictions in these assertions lead one to question their truth. It becomes important, therefore, to test the qualities of

[1] *Fourth Yearbook*, Department of Superintendence, N.E.A., Chapter V, p. 113. Washington: Department of Superintendence, 1926.
[2] A. I. Gates and H. Brown, "Experimental Comparisons of Print-script and Cursive Writing," *Journal of Educational Research*, XX (June, 1929), 14.

these variant forms of letters by an experimental analysis.

3. *Economy of Time for Writer and Reader.*—Communication of thought is the function of the handwriting process. Speed and ease of writing contribute to the economy of time for the writer, and legibility for the reader.

Purpose of the Investigation

The purpose of this study is to analyze certain types of capital letter forms with a view to discovering those basic elements of form which make for speed, stability, and legibility of writing. Certain specific questions are involved:

1. How do three forms of each of four capital letters compare as to speed and stability?
2. How do three forms of each of the capital letters compare as to legibility?
3. What generalizations relative to speed, stability, and legibility can be made regarding the forms of this study?

Speed and stability were measured by means of a test wherein the same capital letter was written for two minutes.

Stability is a term which refers to the state or quality of fixedness of the shape or form of the letter. If instability characterizes the letter forms in any given piece of writing, the condition may be manifest in the deterioration or in the improvement in quality of writing. Deterioration means that the quality grows worse as the writing proceeds. Improvement means that the quality gets better as the writing proceeds. The degree of stability for letter forms can be measured by having competent individuals judge the writing by the general impression which the letter forms make as the writing proceeds. This method was employed in this study.

Legibility was measured by the reports of the perceptual reactions of individuals who were exposed to images of letter forms.

Methods of the Investigation

Speed and Stability.—The appraisal of letter forms with respect to speed and stability was based on a study of three forms of each of four capital letters. This part of the study was made in the Alfred High School, Alfred, New York. In this school the Palmer system of handwriting was taught in the elementary grades.

The selection of the four capital letters was determined by the principle of "the letters with the greatest number of variant forms". One hundred and sixty-three students, ranging from the fifth grade to the senior year in college, submitted variations for each letter. The result was that the letters F, G, H, and T most closely approximated the foregoing principle. For each of these letters three general types of forms were found sufficiently different in style to appear to make possible differences in speed and quality of production. Accordingly, the following letters and their corresponding type forms were chosen for this experiment:

	Type forms		
Letter	I	II	III
F			
G			
H			
T			

Three groups of subjects, equated as to speed and quality of writing, one for each type-form of a letter, were employed in this investigation. An equal number of good and poor writers were included in each group. The average speed score (Ayres' Gettysburg edition) for the good writers was 120 letters per minute, and for the poor writers the score was 97. The average quality score for the good writers was 70, and the score for the poor writers was 36. Six subjects, three poor and three good, were chosen for each group, making a total of 18.

A training period was provided to equalize the differences in familiarity on the part of the subjects with certain letter forms. The training period took place from May 10 to June 14, 1929, during which time the writer and his assistants instructed the subjects in the practice of the various forms. Each group practiced one letter form three periods of 27 minutes each.

The measure of speed was the number of letter forms written per minute.

A measure of stability was achieved as follows: The first ten letters and the last ten letters[3] written during a two-minute period were compared as to quality by three judges[4] according to the following chance arrangement:

19	17	11	16	13	18	20	15	14	12
5	9	2	7	4	1	8	3	10	6

The judges compared each vertical pair of letters, deriving stability scores in terms of units of change according to grades of superiority[5] by the use of + 3, + 2, + 1, and zero.

Stability scores show degrees of change from no change to a possible ± 90. This change may be on the side of improvement or on the side of deterioration in quality. A plus sign indicates improvement in quality in that the score for the first ten is less than the score for the last ten letters of the test. A minus sign indicates deterioration in quality in that the score for the first ten is greater than the score for the last ten letters of the test. Scores showing little or no change indicate stability, and scores showing a relatively large amount of change indicate instability.

Legibility.—Three questions are involved in the analysis of legibility.

1. How do three forms of each of the letters F, G, H, and T compare as to legibility?
2. How do eighty-one forms for all of the letters of the alphabet compare as to legibility?
3. What are the elements which make for legibility?

The three forms of each of the letters F, G, H, and T are the same ones used in the preliminary study of speed and stability. A comparison of these forms as to speed, stability, and legibility is thus made possible.

The eighty-one forms for all of the letters of the alphabet were selected on the basis of variations and representativeness of forms as determined by frequency of use and as determined by types of forms found in handwriting systems.

Students ranging from the fifth grade to the senior year in college were asked to list for each letter of the alphabet those forms which they used in their daily writing. Another list of the forms found in the Palmer, Mills, MacLean, Bennett, and Barnett systems was compiled. Thus a wide variation of forms for each letter was found.

The final selection was made by selecting three forms[6] for each letter of the alphabet. These forms were culled out in such a way that they differed from one another to a considerable extent. Figure 2[7] shows the three forms selected for each of the letters of the alphabet. The forms are classified into the following types: Print, Loop, Hook, Underswing, Semi-print, and Miscellaneous.

The distribution of the forms selected is as follows:

1. Three forms for each of	21	letters -- 63
2. Four forms for each of	4	letters -- 16
3. Two forms for	1	letter -- 2
Total	26	letters, 81 forms

Generalizations concerning legibility are made on the basis of eighty-one forms.

The experimental procedure was that of employing groups of subjects for a given period of time for the purposes herein described.

The subjects were taken from two high schools and a university. The distribution is as follows:

Section A—For 12 forms of the letters F, G, H, T:

1. Alfred University[8] sophomores	30
2. Alfred University seniors	30
3. Alfred University juniors	44

Section B—For 69 forms of all the letters of the alphabet:

4. Alfred University sophomores	52
5. Alfred High School[9] students	31
6. Hornell High School[10] students	45
Total	232

[3] Represents 20 letters out of a possible 40 to 100 letters.
[4] Three female college students.
[5] By superiority is meant: approximation of the written form to the model form; superiority in shape, line, symmetry, legibility, and general merit.
[6] Certain forms represent a large number of minor variants. Three forms, generally, were found to represent these variants for each letter.
[7] See figure near the end of this report.
[8] A university of 600 students, Alfred, N. Y.
[9] A high school of 100 students, Alfred, N. Y.
[10] A high school of 1,600 students, Hornell, N. Y.

The experimental procedure in this study required the use of apparatus to measure legibility. A short exposure method, adapted for class use, seemed superior for the purposes of this study in that many subjects could be used at one sitting. In order to make possible the exposure of characters to a number of subjects at a time, an apparatus was made by the writer whereby letters, copied on slides, were exposed on a screen for a fraction of a second. The materials of the experiment consisted of a Spencer Delineascope projecting lantern, slides, an Eastman Kodak, an enclosed box, a translucent screen, and sheet forms for the recording of reactions. The projecting lantern employed a 500-watt, 115 volt Mazda lamp. The dark box and translucent screen were used in order that the whole mechanism could be set in front of the subjects. This procedure reduced the amount of distraction due to light waves from the projecting lantern.

The Kodak provided the timing device for the experiment. The question of the proper amount of time for each exposure was answered by trying out 1/100 second, 1/50 second, and 1/25 second. It was found that 1/25 second provided the best conditions for the experiment. The forms, written on slides, were exposed for 1/25 second, which occurred one second after the "ready" signal. A few moments pause transpired between the slide just presented and the next one to follow. The forms were presented in random order.

Attention was given to the following general conditions: Instructions were given to the subjects; the light spot on the screen was marked off; the window shades were pulled down just enough to provide clear images to be perceived and enough to provide light for writing down reactions; and moving and talking were minimized as much as possible.

Results of the Investigation

A. *Speed, Stability, and Legibility*

Speed, stability, and legibility are the characteristics of letter forms used in this study. Three forms of each of the capital letters F, G, H, and T are employed as a basis for comparison and analysis. Speed is measured in terms of letter forms written per minute. Stability is measured in terms of units of change, from no change to ± 90. Legibility is measured in terms of the number of correct responses to the exposed forms.

Table I shows a comparison of the forms with respect to these characteristics.

Table I

Comparison of Certain Capital Letter Forms With Respect to Speed, Stability, and Legibility*

	Form	Speed	Stability	Legibility
1.		40	+ 6	73
2.		25	—17	27
3.		31	+15	58
4.		50.5	+ 3	72
5.		38.3	+ 1	88
6.		46.9	— 5	58
7.		37	— 3	82
8.		34.3	— 5	67
9.		38	0	86
10.		47	—10	86
11.		37	—14	65
12.		55.3	— 2	77

* Speed and stability scores based on records of 18 subjects, legibility scores on 104.

For the letter F, form 1 is considerably superior to forms 2 and 3 in the three characteristics measured. Form 2 shows the poorest record. For the letter G, forms 4 and 5 are about equally superior. The relatively poor showing of form 5 in speed, however, would give form 4 a slight advantage. For the letter H, the differences are relatively small. Forms 7 and 9 are about equally acceptable on the basis of the three characteristics measured. Form 9, however, appears slightly superior in stability and in legibility. Form 8 is probably the poorest. For the letter T, form 12 seems to have a slight advantage over form 10, particularly in speed and stability. Form 11 is considerably inferior to both of these forms.

Further inspection of Table I reveals the fact that forms are generally uniformly superior or inferior in all three characteristics measured. It is possible, then, to select letter forms on the basis of speed, stability, and legibility.

An analysis of the best and poorest forms with respect to speed, stability, and legibility contributes to certain generalizations.

1. *Forms should have relatively few parts or lines.* Form 1, for example, has fewer parts than does form 2. Form 12 has fewer lines than form 11. Forms 1 and 12 are superior forms.

2. *Forms should have no double-rocked lines.* Forms 2 and 11, for example, involve double-rocked lines, and are slow, unstable, and relatively illegible.

3. *Filigree work and complex parts within forms should be avoided.* Forms 2, 3, and 11 are particularly encumbered with filigree work, and are slow, unstable, and illegible.

4. *Straight lines should be used wherever possible, particularly in place of double-rocked lines.* This generalization is especially applicable to lines in forms 2, 10, 11.

5. *Underswings at the beginning of the formation of the letter decrease the amount of speed, stability, and legibility.* Form 8 of the letter H is reduced in speed by four letters as compared with form 9. The difference is apparently in the time taken to write the initial underswing in form 8. Form 5 of the letter G and form 10 of letter T are also affected by the time taken in writing the underswings.

6. *Square and triangular forms aid in legibility but not necessarily in speed and stability.* Form 5 of the letter G, a square type, shows a good legibility score and a poor speed score. Form 10 of the letter T shows a good legibility score and a relatively poor speed and stability score. This form is somewhat triangular in shape.

7. *Continuous lines, unbroken by a succession of angular changes in direction, aid speed, stability, and legibility.* Forms 1 and 12 illustrate this principle. Form 2 is an example of a form broken by a succession of angular changes in direction.

B. *Legibility*

Eighty-one letter forms, generally three for each letter of the alphabet, were studied with respect to legibility by the use of tachistoscopic methods. Two hundred and thirty-two subjects, from high school and college, participated in the study.

The eighty-one letter forms are shown in Figure 2. The three or four type forms for each letter are classified as Print, Loop, Hook, Underswing, Semi-print, and Miscellaneous.

The *most legible, medial legible,* and *least legible forms* for each of the letters of the alphabet are indicated by the * mark, no mark, and the # mark respectively. Print forms, with the exception of form 20, are most legible. Hook forms and underswing forms are particularly illegible. Loop forms show a fair record.

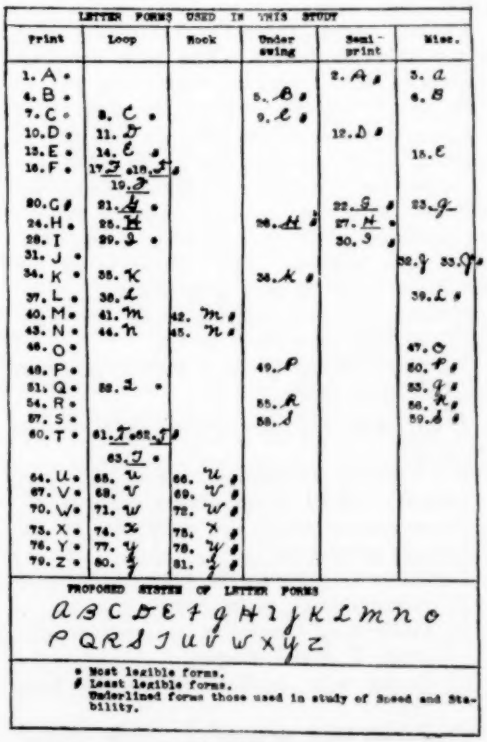


Figure 2.—Letter Forms Studied with Respect to Legibility.

The letter forms in *order of legibility* are as follows: 70, 43, 1, 57, 67, 16, 79, 24, 46, 7, 13, 60, 3, 8, 54, 73, 48, 21, 27, 61, 64, 51, 34, 52, 76, 25, 10, 63, 37, 40, 4, 29, 31, 11, 17, 23, 6, 32, 35, 47, 26, 29, 62, 77, 38, 58, 68, 78, 59, 44, 41, 19, 22, 65, 55, 33, 49, 71, 20, 56, 15, 36, 66, 74, 5, 72, 30, 39, 14, 45, 9, 12, 50, 2, 75, 80, 42, 53, 18, 69, 81. The average score (number of right responses) for all of the twenty-six print types was 104; for the twenty-three loop types, 78; for the eight hook types, 41; and for the seven underswing types, 61.

The seven most legible forms in the above series are print forms for the letters W, N, A, S, V, F, and Z. These forms are generally of the shape of squares, rectangles, and triangles. Lines are straight, except as found in form S.

Although print types of forms are found to be most legible, the case for the cursive type is not a hopeless one. Cursive forms 3, 8, 21, 27, 61, 52, and 25 show scores approximately equal to the average score of legibility for print types, (104).

Certain letter forms are confused with forms of other letters. The most frequent confusions are: form 53 with form 23; 81 with 78 and 17; 80 with 23, 52, and 17; 9 with 21, 22, and 38; 14 with 57 and 59; 71 with 65 and 25; 38 with 3; 42 with 44; 18 with 62. Analysis of these confusions together with introspective reports reveals the fact that similarity of forms is the chief cause of confusion. The "filling in of the perceptual gap" is also a common cause of confusion. This is illustrated in the confusion of form 62 with form 18.

From the foregoing facts, certain generalizations can be made.

1. *Print forms are more legible than cursive forms.*
2. *The order of legibility according to type of form is: Print, Loop, Underswing, Hook.*
3. *Some cursive forms are as legible as the average print form.*
4. *Legibility is aided by the following factors:*
   a) Straight lines, as in forms 43, 1, 79.
   b) Outlines of triangular, rectangular, circular, T, and L shapes.
   c) Forms with outlines above the base line.
   d) Forms with straight or simple curved lines (not double-rocked) with much white space, as in forms 3, 27, 63.
   e) Forms wherein white space is not obliterated by curves, loops, and filigree work.
   f) Forms not too similar to forms of other letters, as 1, 79, 24.
   g) No similar loops within a form, as in 21, but not in 80; no hooks and curves moving in many directions, as in 67, but not in 69; no underswing approach strokes in combination, as in 6, but not in 5.
   h) Distinct features and differentiating parts are helpful in perception, as in 61.

Conclusions

Great variations in speed, stability, and legibility exist in handwriting. Speed scores range from 25 to 50 letters per minute. Stability scores range from no units of change to −17, (out of a possible ± 90). Legibility scores range from 5 to 118 correct recognitions, (out of a possible 128 correct recognitions). Forms and their confusions range from no confusions to 55.

Stability, a characteristic given but little attention, is important in writing. In the selection of letter forms, this characteristic should be considered because of the need for uniformity in writing. Stability can be measured by reliable measuring instruments.

There is a need to select letter forms with respect to speed, stability, and legibility. Some forms should be eliminated from handwriting. Other forms should be greatly modified.

Forms can be selected on the basis of *speed, stability,* and *legibility.* Underlying principles and factual data support this statement.

Although print types of letters are most legible, some cursive forms are as legible as many print forms. These cursive forms are generally of the loop type. Hook and underswing types are so low in legibility as to warrant their elimination from handwriting systems.

A new alphabet of capital letters, based on the facts and principles herein found, is presented in Figure 2.

IN THE USEFUL FUNDAMENTALS OF ARITHMETIC IS IT POSSIBLE TO OVERCOME TEMPORARY LAPSES IN ABILITY LEADING TO STEADY INCREASE IN ERROR?

GUY M. WILSON
WITH THE COOPERATION OF
EVERETT R. PORTER
Boston University

In a recent scholarly article, Knight and Ford[1] report data supporting these theses:

(1) "Multiplying a complete multiplicand by any digit . . . becomes more difficult the further along in the total work it appears."

(2) "The later a multiplication fact appears in a series, the more frequent the error upon it, even when its associations with necessary carrying are made negligible factors".

The above conclusions are so disturbing that they should not be permitted to go unchecked. This study is a preliminary attempt at checking.

Knight and Ford secured their data by administering four test forms to 200 pupils in the sixth grade of four schools. One hundred fifty-two random selections of units of work were critically examined. The product positions which contained the 163 most frequent errors were examined critically.

I reflected upon this article. My experience years ago with traditional arithmetic would have led me to accept its conclusions. My experience the last few years with the co-operative experiment for 100% arithmetic, led me to doubt the necessity of such results as those reported by Knight and Ford.

In other words, my experiment with the 100% plan in the fundamentals of arithmetic, based upon proper motivation and understanding, and a more perfect drill service, is leading me to doubt the necessity of more or less constant error on the part of pupils due either to inadequate mastery, so-called "carelessness", or the lapse of time in a recitation period. Acceptance of such conclusions as those set forth by Knight and Ford is an admission of failure on one of the simplest and easiest tasks of the schools—the mastery of the fundamental tool materials of basic subjects. Such failure should not be taken for granted; it should become a matter of concern.

Accordingly, I invited in two of my co-operators and planned, with their assistance, to test the disturbing conclusions of Knight and Ford. The brevity of the report of the experiment by Knight and Ford made it difficut to re-construct the test or to know just what details were critically studied. Following as best I could, and deciding to pay particular attention to the multiplication fact, 7x9, as did Knight and Ford, I prepared the following test.

To the Children:

Do this test at your usual rate. Check as you go.

When through, turn paper over on your desk and take up other work until all finish.

Multiply:

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
946	649	496	694	469	946	964	649
x73	x37	x713	x73	x73	x37	x73	x73

This test was sent to co-operators in Revere and Lynn, Mass.,[2] with the request that it be given to children who had the 100% ideal, but who had not yet reached the goal. It was accordingly given to lower I.Q. sections in sixth and fifth grades, and to some fourth grade pupils who were in the midst of mastering multiplication.

The papers returned numbered 370. Of these, 198 showed perfect scores. Of the 172 papers showing mistakes, only 65 pupils made

[1] F. B. Knight and E. Ford, "Temporary Lapses in Ability and Error in Arithmetic," *Elementary School Journal*, XXXII (October, 1931), 111–24.

[2] To Rosa A. Yeomans and Helen J. Piper respectively.

an error in handling 7x9. These 65 pupils made 125 errors on the 7x9 facts. The results are summarized by grades and by totals in Tables I and II.

Table I is arranged to show the errors on the 7x9 combinations in the different positions, from first to eighth. Figure I shows the same data graphically. It is evident that these children do not show "unvaryingly an increasing tendency toward inattention and failure" as they move from position one to position eight in meeting the fact 7x9 in examples.

TABLE I

ERRORS IN THE 7x9 COMBINATION FOR GRADES 4, 5 AND 6, ARRANGED TO SHOW THE RELATION OF POSITION OF THE COMBINATION IN THE EXAMPLE TO FAILURE, IF THERE IS A RELATION

Position	Fourth Grade 38 pupils 9 errors	Fifth Grade 184 pupils 80 errors	Sixth Grade 148 pupils 36 errors	Total 370 pupils 125 errors
1 (2d example)	1	7	1	9
3 (6th example)	0	16	11	27
4_1 (5th example)	2	7	2	11
4_2 (8th example)	1	2	4	7
2 (4th example)	2	12	2	16
6_1 (1st example)	0	12	8	20
6_2 (7th example)	0	12	5	17
8 (3d example)	3	12	3	18
Total	9	80	36	125

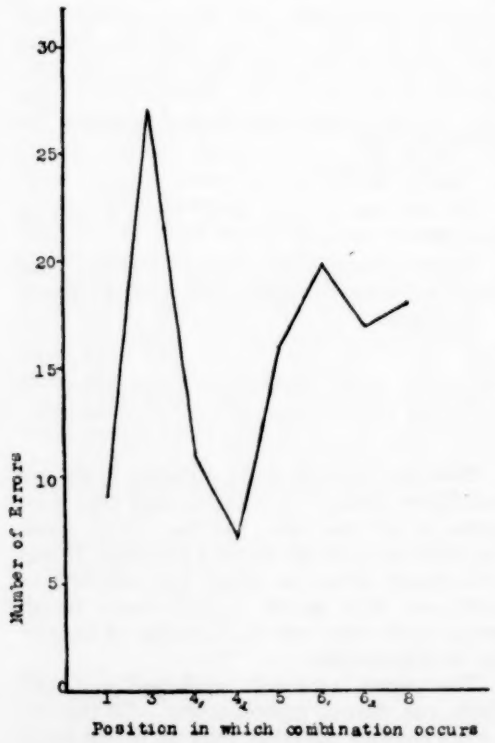


FIGURE 1.—DISTRIBUTION OF ERRORS IN "7x9" IN ALL GRADES BY PLACE ORDER (370 PUPILS, 125 ERRORS)

Let us explain the meaning of "position" as here used, in order to make the meaning of Table I and the above paragraph clearer. In the example 45x32, the 2x5 product figure

4 5
3 2
——
2 1
4 3
——

comes in position 1, the 2x4 in position 2, the 3x5 in position 3, the 3x4 in position 4, and so on for any number of positions in a larger example. The position of the 7x9 combination in the test occurs as follows:

1st example, 6th position
2d " , 1st "
3d " , 8th "
4th " , 5th "
5th " , 4th "
6th " , 3d "
7th " , 6th "
8th " , 4th "

Table II is arranged to show the effect of lateness in the series of examples. Figure 2 shows totals graphically for this table. Again it is evident that these children do not show "unvaryingly an increasing tendency toward inattention and failure as they progress through the exercise from first example to the last one."

TABLE II

ERRORS IN THE 7x9 COMBINATION FOR GRADES 4, 5 AND 6, ARRANGED TO SHOW THE EFFECT OF LATENESS IN THE SERIES OF EXAMPLES, IF THERE IS AN EFFECT

Order of Occurrence	Positions	4th Grade 38 pupils 9 errors	5th Grade 184 pupils 80 errors	6th Grade 148 pupils 36 errors	Total 370 pupils 125 errors
Example 1	6	0	12	8	20
Example 2	1	1	7	1	9
Example 3	8	3	12	3	18
Example 4	5	2	12	2	16
Example 5	4	2	7	2	11
Example 6	3	0	16	11	27
Example 7	6	0	12	5	17
Example 8	4	1	2	4	7
Total		9	80	36	125

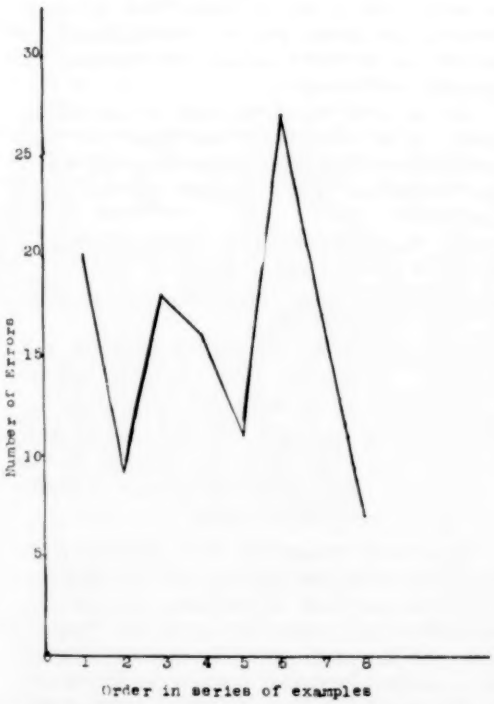


FIGURE 2.—DISTRIBUTION OF ERRORS IN "7x9" IN ALL GRADES ACCORDING TO LOCATION IN SERIES OF EXAMPLES (370 PUPILS, 125 ERRORS)

SUMMARY AND TENTATIVE CONCLUSIONS

Knight and Ford gave their test to 200 sixth grade pupils; they selected 152 random units of work for analysis; 163 errors were critically studied. My test was given to 370 pupils in the fourth grade and in weaker sections of fifth and sixth grades. The 7x9 combination was critically studied, following Knight and Ford. There were 125 errors on this combination, made by 65 pupils. There were 198 of the 370 papers that showed perfect scores. All of these pupils had been exposed to the 100% ideal.

Our data as summarized in Table I and Figure 1 do not support the Knight–Ford thesis that a combination is missed more and more often as its position in the example comes later and later. In our test, position 3 of example 6 was most missed; position 4 of example 8 was least missed. Position 6 of example 1 was second in difficulty. Obviously, order of position in an example does not establish a rule for order of difficulty as shown by frequency of errors.

More careful study of Table I does show the easiest situation, and that is the situation that does not involve any carrying. Thus, we find that the three points of lowest error are position 4 of the 8th example, position 1 of example 2, and position 4 of example 5. Each of these is the first multiplication for a partial product, and so does not require that anything be "added in" before the answer is "set down". On the other hand, errors are practically equal in positions 5, 6 and 8.

Table II and Figure 2, showing the relation of the serial number of an example to the errors made in a combination, do not support the thesis that lateness in the series increases the number of errors. Example 6 shows most errors in the 7x9 combination, example 1 comes second, example 3 comes third, and

example 7 comes fourth in number of errors; while the 8th example, the last in the series, shows fewest errors, due no doubt to the absence of carrying, as explained above.

How can this contradiction of results between our data and the Knight–Ford data be explained? It may be accidental, since the data in each study are very few and obviously inadequate for a conclusion. I offer my data for publication only to keep the question open. Already I have seen complacent quotations from the Knight–Ford article. In my opinion the indictment is too serious to permit of complacent acceptance.

Another possible explanation is that traditional arithmetic as traditionally taught does lead to inattention and failure, and that therefore the further along a combination is in the example or in the series, the greater the chances of error. Hence, the Knight–Ford data, if drawn from a traditional situation![3]

On the other hand, my data were secured under the conditions of thorough motivation, careful teaching with a step-by-step drill service, and the stimulus of the 100% ideal. Inattention and failure did not develop progressively. Even though less capable and lower grade pupils were chosen, 198 of the 370 pupils turned in perfect papers.

The Knight–Ford study will have served a good purpose if it calls attention to the effects on attention and success of the traditional arithmetic. There has been much advocacy of traditional arithmetic in the last few years and much criticism of the attempts to limit the arithmetic in the grades to adult usage.[4] Some have gone so far as to state in effect that "if certain fractions did not occur at all in life we would still teach them in order to extend the knowledge and understanding of fractions." In general, the traditionalists have supported the notion that arithmetic should be chosen on the basis of "mind training and the development of powers", instead of "usefulness in business and life". The effect upon children has been questionable.

By traditional teaching we mean—but traditional teaching is hard to characterize in a single sentence, and anyway some teaching that purports to be quite modern misses the main point of good teaching, viz., careful adjustment to child needs and interests. In general, teaching in arithmetic is poor if it sets an attainment goal at less than 100% for fundamental drill; if it fails to teach a process in simple steps according to child needs; if it carries the teaching steps quite beyond pupils' needs, e. g., 54 unit skills in simple addition, 200 unit skills in addition of fractions, etc.; or, in sum, if it fails to recognize pupil needs and proper motivation as the core of good teaching.

Is it to be concluded from Knight and Ford, that the traditional teaching of the traditional arithmetic leads inevitably and unvaryingly in the direction of inattention and failure as regular habits? If so, then no doubt there are disciplinary effects which are unsatisfactory even to those who advocate mental discipline and the development of powers as the chief reasons for teaching elementary mathematics.

On the other hand, the data of the present study, while admittedly insufficient for final conclusions, do suggest trends in a very different direction. They suggest the following possibility:

Given, —simple teaching methods adjusted to child learning
—meaning first, thus avoiding confusions
—omission of useless processes and details, thus avoiding annoyances and quick forgetting.

We find,—no indication of the oncoming of inattention and failure as the work progresses through a study-recitation period.

Apparently, therefore, it is appropriate to suggest to a teacher whose pupils are developing inattention and failure, that she carefully scrutinize and revise her goals, her curriculum, and her methods. The work in elementary arithmetic should lead to perfect scores in the simple fundamentals with interest and attention undiminished during an ordinary class period, and it doubtless will if the curriculum in arithmetic is well conceived, the placement of topics properly timed, and the teaching well done.[5]

[3] Some of the examples used in the Knight–Ford test go beyond common adult usage, for example, $\frac{4537829}{7}$ $\frac{47038}{2079}$

[4] For example, *Report of the Society's Committee on Arithmetic,* Chapters I. II. Twenty-ninth Yearbook of the National Society for the Study of Education. Bloomington, Illinois: Public School Publishing Co., 1930.

[5] See G. M. Wilson, "New Standards in Arithmetic," *Journal of Educational Research,* XXII (December, 1930), 351–60.

THE EFFECT OF BILINGUALISM ON THE MEASURED INTELLIGENCE OF ELEMENTARY SCHOOL CHILDREN OF ITALIAN PARENTAGE[1]

HARRY SEGNER HILL
Principal Trenton Public Schools, N. J.

This investigation was an experimental study of elementary school children, who were alike in nationality, school progress, social status, and in certain other factors, and who were unlike in the kind of language spoken within the homes. The primary purpose of the investigation was to weigh the effect of bilingualism on the scores of a variety of recognized tests of intelligence by comparing Italian children who heard and spoke Italian at home with Italian children who heard and spoke English at home.

In planning the study, the writer hoped that some light might be thrown upon the following questions: Will Italian children who hear and speak Italian at home do as well as Italian children who hear and speak English at home on a verbal test of intelligence, on a non-verbal or non-language test, on a performance test, on the various sub-tests of a group test, and on a spelling test? How well does a non-verbal test given the Italian child when first entering school indicate what that child will do later on an extremely verbal test? Does familiarity with the Italian dialect increase with chronological age? Is the more intelligent Italian-speaking child also the more familiar with the Italian dialect? Is a verbal group test as reliable a measure of the intelligence levels of Italian-speaking children of Italian parentage as a non-verbal or non-language group test?

The groups compared with each other in the study were designated as Ie and Ei. Group Ie was composed of those children, whose parents were Italian, and who had always been accustomed to the use of an Italian dialect within the home. Children in this group have been influenced more by the amount of Italian than by the amount of English spoken in the community. Since Italian had always been the language of their homes, their English was of a low order when they first entered school.

Group Ei was composed of those children, whose parents were also Italian, and who had always been accustomed to the use of English at home. Children in this group heard Italian spoken within the community, but since English had been the language of the home, their Italian was of a low degree of efficiency in respect to English. In fact they did not comprehend it at all. When they first entered school, they were already familiar with the language of the school.

The initial step was taken by using the results of the Questionnaire on Italian Language Background and the additional information supplied by the visiting teacher of the school. Children who had indicated by the questionnaire that the Italian dialect was the prevailing language of the home were subjected to a test of the Comprehension of Spoken Italian. Proof that they were thoroughly conversant with the Italian dialect as evidenced by the results of the comprehension test prompted the investigator to administer to these children other tests of the knowledge of Italian. Three tests of Italian Word-Meaning were administered and the composite scores obtained on the three tests were recorded for each case. Certain that here was a group of children who were conversant with the Italian dialect and also recognized commonly used Italian words, the writer checked the list of cases and selected sixth-grade children of Italian parentage whose fathers were common day laborers, who had made the same normal progress through the grades (six years including the

[1] An abstract of a thesis submitted in conformity with the requirements for the degree of Doctor of Education at Rutgers University: bibliography, 157 references; dissertation committee: E. C. Witham, Chairman, C. E. Partch, J. A. Blackburn, J. L. La Poe, R. B. Cunliffe, and D. A. Prescott.

kindergarten), who were the first- or the second-born in their respective families, who were born in Trenton, and had lived in the neighborhood since first attending the same public school in the neighborhood. By restricting the selection of the subjects in this manner, the writer was enabled to place twenty-five boys and twenty-five girls in Group Ie.

These fifty cases in group Ie were paired with fifty cases in Group Ei, who were accustomed to the use of English at home. The pairing of Groups Ie and Ei was made on the basis of sex, chronological age, mental age derived from the National Intelligence Test (Scale A, Form I), and intelligence quotient. Both groups were of the same nationality. Each subject in Group Ei and Ie had spent six years in the same public school, which the subject had attended regularly since first entering the kindergarten. Progress through the grades had been normal and regular for each child. In pairing children on the order of birth, first-born was paired with first-born in all cases save seven, in which second-born was paired with second-born. All these children were born in Trenton, and have lived in the neighborhood all their lives. Their fathers were in the same occupational class, that of the common-day-laborer, and were earning a weekly wage of $12 to $18 at the time this study was in progress. This range represented the wage scale prevailing for the last four years in the large mill in the neighborhood and in other factories in other parts of the city. In Group Ei ten fathers were on the relief rolls of the State Emergency Relief Administration. In Group Ie this number was twelve. The weekly wage reported for these during their last season of employment also fell within the restricted wage range reported above. The rentals paid for homes in the neighborhood ranged from $15 to $20 according to the Federal Census of April, 1930. In 1934 they were lower than $15. A large proportion of dwellings lacked inside facilities. They had no bath-tubs or showers, no gas or electricity, or even water. Along with indications of sub-standard housing, there was also evidence of overcrowding of homes. A Scale of the Socio-Economic Status of the home was used to portray the degree of homogeneity which existed in the homes of this particular social class. Together with the room-person ratio, the socio-economic scale presented the social and economic homogeneity of the two groups under comparison, Group Ei, children of Italian parentage speaking English within the home, and, Group Ie, children of Italian parentage speaking an Italian dialect within the home.

The most prominent feature of their dissimilarity was the ability to use and to understand the Italian dialect of the community. This ability was determined by a Questionnaire, Tests of Comprehension of Spoken Italian, and Tests of Italian Word-Meaning. The mental ages of the children in both groups were determined by the scores obtained on the National Intelligence Test, Scale A, Form I. These groups were then subjected to the Otis Self-Administering Test of Mental Ability, Intermediate Examination, Form A, List 8 of the Morrison–McCall Spelling Scale, the Army Beta Group Examination, and the Witmer Cylinder Test.

On the third-grade level, thirty-six Italian children who heard and spoke Italian at home were matched in chronological age, mental age, intelligence quotient, and sex with thirty-six Italian children who heard and spoke English at home. The paired cases were subjected to the same restrictions of selection as the cases on the sixth-grade level. Mental ages for the third-grade children were determined by means of the Haggerty Intelligence Test, Delta I. The groups on the third-grade level were subjected to the Pintner Non-Language Mental Test, List 8 of the Morrison–McCall Spelling Scale, and the Witmer Cylinder Test.

On the first-grade level, thirty-six Italian children who heard and spoke Italian at home were compared with thirty-six Italian children who heard and spoke English at home. An analysis of the performance of the two groups was made on the basis of language and non-language sub-tests of the Stanford Revision of the Binet–Simon Tests.

Using the raw scores obtained by Groups Ei and Ie on the different scales and tests of intelligence, together with their various sub-tests, the writer computed the following statistical constants: The mean raw score for each group; the standard deviation for each distribution for both groups; the standard error of the means of each test, sub-test, and scale for both groups; the difference of the means of Group Ei and Group Ie in each test, sub-test, and scale; the standard error of

these differences; and the quotient of the difference of the means divided by the standard error of the differences for each test, sub-test, and scale for Groups Ei and Ie.

On sub-tests of the National Intelligence Test (Scale A, Form I), on the Otis Self-Administering Test of Mental Ability (Intermediate Examination, Form A), and on List 8 of the Morrison–McCall Spelling Scale, there were no significant differences between the means of Group Ei and Group Ie. Only on sub-tests 2, 3, 5, and 6 of the Army Group Examination Beta were there significant differences between the means of Group Ei and Group Ie, and of these, the differences on sub-tests 2 and 6 were in favor of Group Ie. On the first and third trial results of the Witmer Cylinder Test, the differences between the means of Group Ei and Ie were fairly significant and in favor of Group Ie, but on the shortest trial, this difference became less significant. On the Otis test 46.8 per cent of Group Ei made scores which were above the mean of Group Ie, and 46.5 per cent of Group Ie made scores below the mean of Group Ei. On List 8 of the Morrison–McCall Spelling Scale, 42.5 per cent of Group Ei made scores above the mean of Group Ie, and 43.3 per cent of Group Ie made scores below the mean of Group Ei. On the Army Group Examination Beta, 49.2 per cent of Group Ie made scores above the mean of Group Ei, and 49.2 per cent of Group Ei made scores below the mean of Group Ie. On the shortest trial results of the Witmer Cylinder Test, 42.5 per cent of Group Ei made scores above the mean of Group Ie, and 38.5 per cent of Group Ie made scores below the mean of Group Ei. The chances were at least 4 out of 10 that a sixth-grade child in one group would do as well as a sixth-grade child in the other group.

Incidental studies by the correlation method revealed that the more intelligent the Italian-speaking child was, the more readily did he acquire the use and the understanding of the English language, and that the more intelligent were also more conversant with the Italian dialect. This familiarity with the dialect increased apparently with chronological age.

On sub-tests of the Haggerty Intelligence Examination (Delta I), and of the Pintner Non-Language Mental Tests, there were no significant differences between the means of Group Ei and Group Ie, except on sub-test 3 of the Pintner Test, on which the difference was in favor of Group Ei. On List 8 of the Morrison–McCall Spelling Scale and on the Witmer Cylinder Test, no great differences between the means of Group Ei and Group Ie were found. On total scores of the Pintner Test, 46.5 per cent of Group Ie made scores above the mean of Group Ei, and 46.1 per cent of Group Ei made scores below the mean of Group Ie. On the shortest trial results of the Witmer Cylinder Test, 47.7 per cent of Group Ei made scores above the mean of Group Ie, and 47.3 per cent of Group Ie made scores below the mean of Group Ei. The chances were at least 4 in 10 that a third-grade child in one group would do as well as a third-grade child in another group.

The test used to measure the intelligence of first-grade groups was the Stanford Revision of the Binet–Simon Tests. Group Ei was superior to Group Ie on those tests involving a moderate understanding and use of the English language, but Group Ie was superior to Group Ei on the tests involving an extremely high degree of efficiency of understanding and use of the English language.

The results of the investigation point to certain conclusions which may be summarized on the basis of the questions suggested in the second paragraph of this report:

1. Italian children who hear and speak Italian at home are not inferior to Italian children who hear and speak English at home on a verbal group test of intelligence. No reliable differences between the groups are found either in the size or the variability of the scores.
2. Italian children who hear and speak Italian at home are not superior to Italian children who hear and speak English at home on a non-language test of intelligence. Fairly reliable differences are found on various sub-tests, but complete reliability of superiority in all the sub-tests is lacking.
3. Italian children who hear and speak Italian at home are only partially superior to Italian children who hear and speak English at home on a test of concrete intelligence. Evidences of a marked superiority are definitely lacking.
4. Italian children who hear and speak Italian at home do as well on the various sub-tests of verbal or non-verbal

group intelligence tests as Italian children who hear and speak English at home.

5. Italian children who hear and speak Italian at home observe, analyze, and recall the parts of English words as well as Italian children who hear and speak English at home.
6. The score obtained by an Italian child on a non-verbal test is a fair indication of the score the same child will obtain later on an extremely verbal test. The coefficient of correlation between the Pintner–Cunningham I.Q.'s and Otis I.Q.'s is .72 ± .038.
7. There is some indication that familiarity with the Italian dialect increases with chronological age. The coefficient of correlation between chronological age and Tests of Italian Word-Meaning is .241 ± .031.
8. There is a small indication that the Italian child who makes a high score on a verbal group test is also more familiar with the Italian dialect of the community. The coefficient of correlation between National I.Q.'s. and a Test of Italian Word-Meaning is .113 ± .054.
9. A verbal group test of intelligence is as reliable a measure of the intelligence levels of Italian children who hear and speak Italian at home as a non-verbal or a non-language group test. Equation of the groups on the basis of a verbal test results in equivalence on the basis of a non-verbal or non-language test.
10. No statistically reliable differences in scores on verbal, non-verbal, and performance tests are revealed between Italian children who speak Italian at home and Italian children who hear and speak English at home.

In the light of the foregoing conclusions based on the data of this investigation, it seems reasonable to believe that the effect of bilingualism on the measured intelligence of Italian children who hear and speak Italian at home may be disregarded.

SYMPATHETIC BEHAVIOR IN YOUNG CHILDREN[1]

LOIS BARCLAY MURPHY
Sarah Lawrence College

This is an exploratory study of sympathetic behavior in nursery school children. The following instruments of research were used to collect data on the same children in different situations:

1. Observation and recording of specific episodes of behavior occurring on the nursery-school playground. The main body of this recorded material was collected by the investigator and her assistants, but additional data were furnished from incidental records of teachers and other observers, and indirectly from the fifteen-minute samples of behavior recorded by investigators studying other aspects of children's behavior such as conflicts and imagination in outdoor play.

2. An inventory of specific behavior items and personality traits. This was rated on a five point scale by teachers of the children in two main research groups twice a year and twenty-four items of it were rated by the investigators each day of their observation.

3. "Framed situations", with animals and children as sympathy stimuli, devised to supplement the situations of the playground; pictures, stories, and questions were also used to compare the children's responses to controlled stimuli.

4. Interviews with parents, providing data regarding the children's behavior at home, experience with pets, younger children and the like, which might influence their sympathetic behavior. These interviews were supplemented by study material from the cumulative records of the nursery-school.

5. Detailed day by day records on six children were kept over a period of a year or more by parents.

The comparison of odd and even items in the observation records gave a correlation of .94 and .99 for the two main research groups. The ratings and observations gave correlations of .7 to .8 calculated separately for the two groups.

The most characteristic stimuli for sympathetic responses among our pre-school children were classified as follows:

SYMPATHY STIMULI

Distress due to:

Physical cause
- Accident
- Attacked by child
- Physical discomfort (cold, uncomfortable position, caught in play apparatus, etc.)

Mental distress
- Toy snatched or threatened
- Play hampered or intruded upon
- Disciplined
- Separated from mother, father, etc.
- Fear

Emotional expression without evidence of stimulus
- Crying
- Holds hands over face
- Pained expression—inhibition of crying
- Evidence of injury without evidence of present pain—sore lip, mercurochrome, bandage, etc.

Wishes, needs, etc.
- Expressed wish (I want __, help me, or some action which reveals want or need)
- Unexpressed wish or need—assumed or implied
- Precarious situation (need for care)

Adult in distress or want
- (Chiefly physical danger or physical need)

[1] This is a partial summary of the influence of group situations and variations in motivation on sympathetic behavior in nursery school children. The material summarized here is part of a larger study of sympathetic behavior in young children. There is no attempt in the present summary to discuss the more fundamental factors of culture and the personality structure of the child as they bear on the understanding of one aspect of social behavior. These are discussed systematically in a forthcoming book, *Sympathetic Behavior in Young Children*. The study was carried on at the Child Development Institute, Teachers College, Columbia University, with the financial assistance of the Josiah Macy, Jr. Foundation.

Animal in distress or want
(Physical danger, need such as hunger, real or supposed attack by another animal or human being)

A similar list of characteristic sympathetic responses gives us the following:

RESPONSES TO SYMPATHY STIMULI

I. *Social or sympathetic responses*
Helps
Helps out of physical distress situation—picks child up **after** falling.
Helps with play—gives toy, plays with child, pushes swing, helps climb.

Removes or attempts to remove cause of distress.
"Stop it!" "Don't!" to aggressor or other child.
Pushes away child attacking another (when this solves difficulty).
Explains child's attitude to adult causing resistance.
Removes cause of physically distressing situation—removes tricycle from fallen child, gets out of child's way, uncatches child's foot.
Stops play when it is in opposition to other child's need—e.g., stops jumping on board while child climbs off.

Comforts.
Pets, pats, hugs, kisses.
Reassures, e.g., "I won't hurt you."
Expression of **solicitude**; e.g., "That's too bad," "That hurts, doesn't it?"

Punishes cause.
Attacks the aggressor of the other child before the aggressor has withdrawn but not thereby remedying the difficulty.
Attacks aggressor of other child after difficulty.
Speaks sternly to aggressor (after difficulty).
Retaliates on aggressor (does to the aggressor what aggressor did to first child).

Protects, defends.
Defense of other child's material.
"It belongs to __", "__ was using that", "You can't have that."
Helps other child hold toy against aggressor.
"Don't hurt __", "Don't touch __" (before difficulty occurs but is imminent).
Retrieves snatched property for other child.
Explains other child's wants to aggressor.
Verbal defense of other child: "He *can* get in." "That's not Jenny __ she's crying."
Protects from hurt—catches child as he falls.

Warns.
Warnings of doubtful motive—look out, watch out, etc.
Sympathetic **warnings**: "You might hurt yourself." "You might fall."

Tells adult or other child.
"Look at __"
"He's crying."
"He wants __."

Questions.
Why is he crying?
Did it hurt you? (to child)
What are you going to do about it? (to teacher).
What's the matter? (to animal or to a child).

Anxious or emotional.
Stares with anxious expression.
Evidence of worry, shakes head, frowns, lips pressed together, etc.
Cries, whimpers.

Suggests or effects solution.
Tells adult to keep other child out of danger.
Tells child how to get out of the situation.

II. *Egocentric and unsympathetic responses*
Laughs.
Scoffs, criticizes.
Laughs, smiles.
Contrasts own behavior ("I don't cry").

Dramatic
Uses situation to play own role, usually dominating.

Sadistic-destructive
Attacks child in distress or his material.
Attacks child who helps child in distress.
Teases, e.g., makes loud noise at.
Ambiguous
Stares.
Comments on situation (usually not addressed to anyone).

Naturally, certain responses are characteristically linked with certain stimuli; physical pain or injury and crying are responded to with staring, then helping or comforting, or both. The staring in response to a cry is sometimes not different from the staring which would be stimulated by any striking sound or sight. Wishes, needs, precarious situations and so forth are responded to with verbal or active help or warning, in case of danger. Whenever distress or pain is the result of an attack by another child, the response of the sympathizer may include punishment of the attacker and defense of the attacked, as well as helping or comforting, although it was more usual for a child to respond to one element in the situation than to all; that is, ordinarily a child comforted the attacked person *or* punished the attacker.

An outline summary of this sort does not convey the subtler aspects of situations like the following:

4/3 Nancy stood by her mother and father and cried.
Beulah stood nearby watching.
Arthur approached, put one arm around Beulah and the other around Nancy. Beulah and Arthur looked worried but said nothing.

3/20 Daniel approached the bus.
Peter pushed him away and pulled his hair.
Daniel squealed.
Wallie watching, shuddered and tensed as Peter pulled Daniel's hair.

5/9 Kirk hit Evan with a rope, continued, hit the teacher as she approached.
Arthur looked on, clenched his fists, shook them, with an intense look. (Emotional reaction to seeing Evan hurt)

3/8 Mary fell and cried.
Douglas and Evan were near; they looked and went off.
Winifred approached, looked, then put her arms around Mary and kissed her.

3/6 (Philomena approached Arthur who had been sitting in the house watching the group through the window.)
Arthur to Philomena: "Is that a white hat?"
Philomena: "Yes."
Arthur: "Is it a warm hat?"
Philomena: "Yes."
Arthur: "I like you. Where is your mother?"
Philomena: "Home."
Arthur: "Where is your daddy?"
Philomena: "Gone to Washington."
Arthur: "Will he come back?"
Philomena said something unintelligible.
Arthur: "Will he come back Christmas time?" (Anxiously with hand on Philomena's shoulder)

1/19 Daniel approached the group.
Nancy pushed him.
Evan hit him.
Daniel knocked Evan down.
Evan cried.
Alice looked anxiously at Evan and patted his face.

10/11 (Reported by staff member):
Kirk walked around crying, "Me have no Mommie."
Agatha followed him, looking very concerned, and put her arm around him.
Agatha asked: "What's the matter, Kirk?" in a solicitous tone.
Agatha turned to the observer and said, in a worried tone, looking serious: "That little boy has no Mommie."

Different ways of estimating the frequency of sympathetic episodes in the two groups intensively studied indicated that on the average one sympathetic response might be expected to occur in an hour's observation of a child, or one-eighth as many sympathetic responses as conflicts for the same group.

This estimate of course is based simply on responses to distress of the sort outlined, and does not include cooperation and friendly behavior of other sorts.

Two groups, "Group H" and "Group W" were observed for two hours on 90 or more mornings each, from October to May. When the number of responses in each group was weighted for number of children present during the period of observation, Group H showed a lead of 26% more sympathetic responses than Group W showed. The two groups were different in organization—Group H included a much wider-age-range, with 30 months difference between the oldest and youngest child, while Group W consisted of 18 to 20 children all within a narrow range of 10 months.

The differences in age in Group H were great enough to prevent constant competition for the same play materials and activities and to stimulate protective, "big brother" types of response of older toward younger children. There may doubtless have been important imitative factors: when older children set examples of protective and sympathetic behavior, these would be copied by younger children when occasions offered. Group W, on the other hand, consisted of children developmentally close enough to be interested largely in the same materials and activities: this necessarily resulted in more competition and conflicts over these materials. And the same imitative factors that were present in Group H, which resulted in younger children picking up cooperative and helpful patterns from the older ones, would result in acquisition of more techniques of conflict in Group W. In this connection it is interesting to note that for three successive years, groups occupying the present quarters of Group W had shown high conflict scores as compared with groups in the quarters of Group H. It looks, further, as if the smaller amount of space (and of properties and play materials) put a heavier competitive pressure upon the children on the one hand and contributed to more frequent physical contact on the other hand, both of which might tend to increase conflicts and unsympathetic responses.

In Group H the division in the schedule of the children tended to make the older children aware of the babies as a definite group. Data from other studies suggest that it is reasonable to emphasize the importance of the age-range for the behavior within the group. Studies of jealousy, for instance, have shown that this type of response between siblings is less likely to occur when there are four years or more of difference between a pair of children than when the children are closer together in age.[1] Closeness in age, of course, creates certain actual competitions through interest in the same toys and materials, and in certain types of response from others in the group; where the difference in age is greater, the competition probably decreases, and along with it, conflict, jealousy, and other aggressive reactions.

Quite aside from the effect of the size and age-range of the group, there were other factors which probably influenced the behavior of the group. We find certain patterns of response to sympathy situations on the part of teachers in Group H and certain other patterns in Group W that are important to consider here. It is difficult to describe these differences in general terms, but the subjective impressions of observers that at least one of the teachers in Group H was outstandingly spontaneous, warm, responsive to the children, and that at least one of the teachers in Group W had certain rather "hard-boiled" patterns is substantiated by records of affectionate comfort in the first instance, and records of repeated laughing at a crying child in the second instance. It is not implied that learning to laugh off one's hurts is not desirable, but simply that a technique, even when it is legitimately and discriminatingly used by an adult with the intention of helping the child, may be picked up by children and used quite indiscriminately and with egocentric motives entirely. It was in Group W that the pattern of contrasting own behavior—usually saying: "I'm a big boy, I don't cry"—spread from one child to another.

Two children who showed this pattern characteristically (using it 7 and 5 times, respectively), as compared with a maximum of 2 times among the other children, were children who had both high aggressiveness ranks (fourth and second in their group of 19) and high intelligence quotients (124 and 123 respectively on the Minnesota pre-school test; it will be remembered that this tends to score

[1] M. Sewall, *Smith College Studies in Social Work*, 1930, I, 6–22.

lower than other tests such as the Kuhlmann.) This suggests that these personalities, unusually intelligent and sensitive to the patterns of adults, and aggressive in their social responses, were pivotal points in what might be seen as a process of cultural diffusion, working down from the teacher to the average children in the group via these particularly responsive and aggressive ones.

In Group H there was a case of influence of a dominant personality in the sympathetic direction. Janet, who had the highest sympathy score among the girls, played frequently with Gwen who showed few sympathetic responses during her first weeks in the group. Toward the end of the spring, Gwen's mother spontaneously commented that Gwen was beginning to "do the sort of thing that Janet did," and the records showed episodes of this sort which had not appeared at all during the preceding four months and which were characteristic of Janet's habitual responses.

Altogether, then, we have evidence for a number of influences contributing to the differences between the behavior of Group W and Group H: the effect of the number of children in relation to space, the differences in the constellation of the groups with particular reference to the age-range, the attitudes and patterns of the teachers toward the younger children, particular patterns of the teachers in relation to sympathy situations, and the diffusion effects of the behavior of certain outstanding children in the group.

Charts (not here reproduced) of the number of sympathetic responses given and the number received throw an interesting light on the place of the individual in the group. For example, Janet, the oldest child in Group H, gives sympathetic responses to most of the children in her group, while she receives sympathetic responses chiefly from the older children. Heidi is predominantly a receiver of sympathetic responses and gets them largely from the older children.

Charts (not here reproduced) of the sympathetic responses of the whole group show differences in the patterns of the two groups. The relatively even distribution of sympathetic responses between girls and boys and the tendency for certain children to be foci for radiation of sympathetic responses in many directions or, conversely, foci for collection of sympathetic responses, are quite noticeable. Children who distribute a few responses among many other children contrast with children who concentrate many responses on a few.

When we look at the responses of all children in each group, there are apparent other differences in the structure of the two groups. In Group W, practically all of the children are actively involved in the group, although there are some isolated children who give almost no responses to other children and receive few also. Group H, on the contrary, shows a definite loading of activity among the older children; the younger ones form a periphery with various degrees of aloofness from the group. Three boys are involved in the group almost solely through other children's responses to them, while two others show some responses to other children in addition to a larger number of responses received from other children. Some children are centers with radii of giving and receiving involving a large number of children, while other children show responses which are comparatively concentrated toward one other child. In some cases a child is a focus for receiving responses and gives few, while in other cases one child it primarily a focus from which responses are given.

Further analysis of the reactions of the children in the group shows a number of things. In Group W there are many more *varieties* or patterns of sympathetic response than there are of unsympathetic response—this may be because there are more stereotyped culture-patterns available for use in aggression and that when you are being nice you have to be more original about it, if you are three years old.

In approximately half the cases in which a given child responds to another child with more than two reactions to distress situations, both sympathetic and unsympathetic responses are given, on different occasions. In one-third of the cases, in this group, a child responds to another child consistently with sympathetic responses, that is, never gives that particular child any unsympathetic reactions, and in one-sixth of the cases, he responds only with unsympathetic reactions, that is, never has a kind word for the particular child involved. It is more usual for a child to get both sympathetic and unsympathetic responses from any other child provided the child reacts to him at all. This recalls certain other recent studies which have shown that among nursery school children,

one fights one's best friends more than children who aren't friends.

It is possible to go further, too, and to analyze individual trends in the kinds of response which different children gave. The most conspicuous example is that of Douglas who gave 43 responses of defense of other children as compared with 10 from the child who gave the next highest number of that particular kind of response. Evan gives chiefly help and defense responses, while Winifred is high for the group on comfort responses such as the following:

4/2/34 Sturgis fell off tricycle.
Julius ran and picked up the tricycle.
Winifred ran and patted and tried to comfort.

3/8/34 Mary fell, and cried.
Douglas and Evan near, looked, went off.
Winifred approached, looked, then put her arms around Mary and kissed her.

Nancy gives a large number of help responses, but also a large number of "anxious" responses which are unusual for the group. These may be called individual trends; a given item of behavior does indeed appear among other children but is unusual with them, so that the child who shows many such responses is completely "off the distribution."

Material from the inventory of items of social behavior was used for analysis of differences related to age, sex, and intelligence, and relations between aspects of social behavior. Sixteen items were rated by each of four teachers in one group and five teachers in another group using a five point scale for each item to indicate the relative frequency of that response. The items of the "sympathy" scale were:

3. Defends rights of smaller child.
5. Tells child not to cry.
10. Defends child attacked by another.
20. Responds to another child in distress by staring with anxious expression.
21. Stops own work to aid another child.
23. Tells teacher or other adult that a child is crying.
25. Attempts to comfort another child with pats, embraces, and the like.
26. Helps a new child get into the group.
29. Responds to another child's distress with disorganized behavior.
30. Inquires from adult cause of another child's crying, in matter-of-fact tone.
35. Inquires from adult cause of another child's crying in anxious tone.
37. Laughs when hears another child laugh.
38. Warns another child of danger.
41. Asks a child if he hurt himself (after a fall or other accident)
42. Asks crying child why he is crying.
44. Helps child out of painful situation.

We find that all measures of intelligence correlate positively with measures of social behavior but that chronological age does not correlate positively with measures of sympathetic or cooperative behavior in the narrow age-range. In the group of children within a ten-months age range, intelligence is more important in relation to social behavior than developmental differences connected with age as such; it is easy to see that a child near the bottom of the range in this group on age, for instance, 40 months old, who had an I.Q. of 150 would have a mental age of 60 months, while a child of 48 months with an I.Q. of 110 would have a mental age of only 53 months. If other developmental differences of a similar sort went along with intelligence, they might easily outweigh age as such. When the difference in age amounts to 24 or 30 months among the children, the oldest of whom are just four, this balancing of age by I.Q. is not so likely.

The 6 items of the Sympathy Scale that correlated highest with age and mental age were as follows (Group H, mean rating of all raters):

Of these the mean rating of Group H was higher than that of Group W on items 21 (stops own work to aid) and on item 44 (helps child out of physically painful situation), which gives added evidence that these behaviors accompany increasing maturity. These comparisons suggest that a high order of "active" response or of "verbal" response is more important than any sequential relation between active and verbal responses taken together; the common element in these verbal and active responses is the *direction* of the response toward release from the distress situation, a direction of response which is not clearly present in such responses as "tells

CORRELATIONS OF ITEMS IN THE SYMPATHETIC BEHAVIOR SCALE WITH AGE AND MENTAL AGE
MEAN RATING OF ALL RATERS IN GROUP H

Item		Correlations (N=18) with C. A.	M. A.
21	Stops own work to aid	.60	.62
38	Warns another child of danger	.76	.77
42	Asks crying child why he is crying	.64	.63
23	Tells teacher that child is crying	.65	.60
41	Asks child if he is hurt	.64	.67
44	Helps child out of physically painful situation	.67	.67

child not to cry", "cries when he sees and hears another child crying", "stares with anxious expression."

Sympathetic behavior scores correlated positively with aggressive behavior scores under all conditions; that is, when scores on sympathetic and aggressive behavior rated by a single individual were compared, when the scores of all raters for each group were combined for the group and when combined ratings (of all raters) for each child in all three groups pooled were used.

Both sympathy and cooperation show higher correlations with age than does aggression in Group H which includes children from two to four. Intelligence measures correlate somewhat higher with sympathy and cooperation than with aggression in this same group, but in the narrower age-range of Group W the correlations of both age and intelligence with each aspect of social behavior are markedly reduced. In the two groups H and W, the different age-ranges show different correlations between the aspects of social behavior measured: cooperation and sympathetic behavior are more highly correlated in Group H, and aggression and sympathetic behavior are not so highly correlated as they are in Group W. This suggests that in Group W the narrower age-range is accompanied by less differentiation among behaviors; that is, we have different forces working—on the one hand a tendency for outgoing behavior of all kinds, both aggressive and sympathetic, to increase with age and, on the other hand, a tendency for older children to approach a stable plateau where sympathetic and cooperative behaviors are emphasized and aggressive behavior decreases as these get the upper hand. If this were true, we would expect to find less relation between aggression and sympathy among older children, although probably in this culture these responses would always be positively correlated to some extent.

An item analysis of relationships both within the scales and between the scales is even more interesting than a comparison of the scales as a whole. The most dramatic relationship which appears here is a correlation of .91 between item 4 (joins attack on one child by another) and item 10 (defends child attacked by another). Item 25 (attempts to comfort another child with pats, embraces, and the like) correlates .58 with item 19 (pushes or pulls another child without regard for his discomfort). Item 44 (helps child out of painful situation) correlates .27 with item 7 (pummels child who falls accidentally). When these item relationships are calculated for the three pooled groups the correlations just given are .69, .34, .36 respectively. In other words, the aggressive and sympathetic uses to which certain acts are put may be less important than the acts themselves. The child may use his fists to defend himself in one situation or to defend someone else in another situation; a grasp or a hug may be aggressive or affectionate with similar differences in situations. It remains to be seen in which children these acts are simply specific "habits" with little relation to the child's personality, in which children the acts are directly satisfying in themselves, and in which children they appear as the surface expression of underlying needs.

The important thing here is not the particular correlation so much as the fact that our analyses, from whatever angle they are approached, show similar results, that is, a tendency for aggressive and cooperative behavior

to be found together in this age range. This tendency, however, in general expresses itself through correlations that are not very high.

To a large extent this fits in with findings of other investigators who found other sorts of social behavior increasing with age during this period, and this quite without regard to the socially approved or disapproved character of the behavior. Sympathetic behavior, leadership, friendships increase hand in hand with resistance, aggressions, and less comfortable behavior of young children. In sum, we may say that during this period both the variety of social responses and amount of social responses normally increase with age and that sympathetic behavior differs in no essential respect from any other cross-section of behavior we might wish to inspect.

It may be more surprising to say that sex differences do not exist and that boys are quite as sympathetic as girls at this period. Probably it is too early for a masculine-dominant culture to make a deep impression of a selective sort; these boys do not yet know that they might some day be considered "sissies" for too ready sympathy, whether sought or given, and extroverted responses even of an aggressive sort are encouraged in girls as well as boys.

The fact that at least from an adult normative or ethical point of view, "antithetical" behavior like aggression and sympathy appears to marked degrees in the same child, makes it imperative that we do not leave the data at the point of the usual statistical analysis, but that we attempt to throw what light we can on the conditions under which variations from one sort of behavior to the other occur.

In a culture as complex as this, similar behavior may appear with quite different meanings depending upon the total orientation and inner state of the child. The "defense of a child who is attacked" occurs in different settings:

(1) A child who is seldom spontaneously aggressive occasionally "punishes" a child attacking another child (for instance, Janet who had a consistently low score in aggressiveness by all measurements, pounded Lucinda when she pushed Davis off a wagon.)

(2) A child who seldom attacks another spontaneously makes a habit of hitting or "spanking" children who have attacked another child.

(3) A child who is frequently aggressive finds defenses the most natural method of expressing sympathy.

In cases like the first, a child who actually avoids physical combat in general, as physically sensitive children frequently do, sometimes resorts to it under special provocation, which may be either social in direction or self-defensive. In the second instance, a child may have considerable aggressive drive, which is effectively restrained by home taboos as far as spontaneous attack is concerned; sometimes in these situations, the opportunity for physical punishment as a legitimate outlet of aggression comes as a release from severe repression. In cases like the third, where a child's thresholds are low for any aggressive response, both spontaneous hits and defenses of others come easily and probably have little difference in value for the children.

Other variations in behavior were associated with shifts in physical need or emotional security. Children like Sturgis and Janet, who were almost invariably sympathetic to other children's wishes and distress as long as no conflict with their own needs was involved, would, however, defend their own play arrangements or property when another child interfered with these. In the case of Julius and Allen, a sense of guilt for either intentional or unintentional hurt to another child resulted in flight from the situation. In certain cases, pressure of the group appeared to be partly responsible for a child's response to the distress of another child. Frustration in an initial friendly approach was followed by an aggressive response even among the youngest children. This shift from a mild, or friendly or affectionate approach to a sadistic one is in fact rather common and recognition of this is very important for an understanding of the fundamental ambivalence of aggressive drives. The mother who gets a dog for her son who teases the neighbor's younger boy too much, and the teacher, who asks her behavior-problem spit-ball firing pupil to take responsibility for the blackboards or the turtles, recognizes this double potentiality in what both the mother and the teacher would call a "need for an outlet." Yet it is implicitly ignored when we think and

speak in terms of aggressive "traits", as though they were inherently destructive.

One of the most interesting illustrations of the possibility of a quick shift of attitude and expression in reaction to a shift in ego-status appeared in a sad drama of hurt egos where Janet's usual consistent friendliness and sympathy turned to aggressive retaliation after she felt she had been intentionally injured.

5/24 Heidi threw a block on Janet's leg (perhaps accidentally).
Janet frowned, then cried.
Heidi laughed.
Lucinda frowned.
Janet said, "Heidi, you hurt me."
Heidi continued to laugh.
Janet said, "How would you like it if I threw that block on your knee, huh, huh?"
Heidi thought Janet was fooling and continued to laugh.
Janet threw Heidi onto the blocks.
Heidi said, "Stop it" and cried.
Janet laughed.
Patrick said, "What's the matter?"
Lucinda looked anxiously in Heidi's face.
Heidi and Janet cried (Janet—sympathetic crying)
Janet tickled Heidi and waved her hands in her face, trying to divert Heidi's attention.
Heidi said "Stop that."
Lucinda said, "Let's go down the slide."
Janet walked away reluctantly with Lucinda.
Patrick said, "What's the matter, Heidi?"
Heidi said, "Janet pushed me down."
Julius watched, worried.
Seth said, "What's the matter?"
Heidi said, "Janet pushed me down."
Seth said, "Teacher, Heidi got hurt."
Patrick patted Heidi's head and followed her.
Joyce approached and stared.
Patrick said, "Janet, did you push Heidi down, did you, Janet?"
Lucinda said, "Yes she did."
Janet said, "Well, she hurt me first. She hurt my knee."
Joyce said, "Why you crying, Heidi, why you crying?"
Patrick said, "Get away, Joyce." Then, "Let her alone", and patted Heidi's face and head.
Joyce said, "Why is Heidi crying?"
Lucinda and Joyce watched from far off.
Joyce's nurse wiped Heidi's face.
Julius approached and watched while the nurse wiped her face.
Reinhardt approached with a ball (no attention to crying).
Julius pulled the ball from Reinhardt. (Julius wanted it for Heidi)
Reinhardt pulled.
Heidi said, "It was Reinhardt's, Julius."
Julius released the ball and said, "Why did you throw the blocks at Janet?"
Heidi said, "Cause I wanted to."
Julius said, "You mustn't."
Heidi said, "Janet knocked me down."
Julius said, "Well, you mustn't."
Heidi ran to Janet and Lucinda and they continued playing with blocks.

Janet was a child with a very high score on observation and on teachers' ratings for sympathetic behavior. Her aggression scores were low. Heidi was a friend, a gentle and younger child whom she had often protected and helped. Yet in this situation, Heidi's persistent misunderstanding of Janet's physical injury and hurt feelings resulted in retaliation on Janet's part and laughter at the injury she had done to Heidi. There is a moment of attempted recompense through a gesture toward distraction of Heidi which is not carried through when Janet's companion suggests a new activity. Shortly after, in answer to comment, Janet justifies herself on the ground that the initial aggression was Heidi's. The whole situation is resolved when Janet's good feeling is reinstated by other children's acceptance of the basis of her retaliation, and Heidi's ego is comforted after a great deal of attention from Patrick, Seth, Julius, and Joyce, so that Janet, Heidi and Lucinda finally go off to play together.

In the instances just cited, sympathetic responses which ordinarily appear in certain situations are inhibited when the ego of the potentially responsive child is threatened. Variation in the security of the child is in fact one of the most important sources of variation in sympathetic response, and results in different shifts from one child to another.[1] For instance, Anthony, a somewhat fearful child, is more sympathetic when he is insecure and less so when he is secure and happy in his group; Julius is less sympathetic, more aggressive and defensive, when he is insecure, while his normal ratio of responses during a well-adjusted happy period includes high sympathetic behavior scores; Arthur is more sympathetic during the early period of his records when he is insecure, but later when he develops more aggressive outlets he becomes more aggressive and less sympathetic when he is insecure.

The shift from sympathetic to unsympathetic orientation toward one's neighbor is tied up with fundamental patterns of personality of the individual. The basic features for our purpose concern the individual's attitudes of group-dependence as compared with group-dominance: the individual who temporarily or over a long period of time is in a position of seeking approval of the group he confronts is apt to have lowered thresholds for sympathetic responses to members of that group, when the individual who is, through dominance or lack of interest, in no such relation to the group, may show no such tendency. The individual who has been relatively sure of his dominance on the other hand may only fight harder in self-defense when his security is temporarily or for the first time threatened; a longer period of insecurity might, nevertheless, shift his orientation to one of attempting to win friendly or sympathetic response through giving it.

Other variations in behavior may be understood from inspection of sequential records of individual children over longer and shorter periods. A specific event may raise the security or satisfaction level in a given child in such a way as to be followed by specific sympathetic or cooperation behavior which had not appeared before. When Stephen, a frustrated young artist, was allowed to make a large painting at school he came home and began spontaneously to help his baby sister. On earlier occasions defeat had preceded sadistic teasing of a neighbor's cat. It is doubtful whether either of these sequences can be interpreted best in terms of compensation; it looks as though being happy and satisfied irradiated into friendly feeling which took expression in helpfulness toward his sister, while frustration irradiated in a similar way into antagonism of which the cat was the target.

The development of a friendship in a child who has been on the periphery of the group may introduce sympathetic responses where they were absent before this friendship. Saul was not a child who showed very frequent active sympathetic responses to other children. Neither had he had very satisfactory relations with other children in the group during the first half of the year. But shortly after Gwen came, these two developed a friendship which lasted for a couple of months and which was a steppingstone to more constructive and varied social responses for both children. During this period, when Gwen was hurt, Saul responded to his friend with vivid comfort. In this case, behavior which Saul had never shown before was released by the response to his new friend. It is possible that both psychological and neurological considerations would have to be invoked here for a complete explanation of Saul's response; that is, rubbing a bump is one of many patterns of response that may be found together in an experience of friendship—it is one of the things you can do for a friend (if you are four years old) and also, the positive emotional elements present in a friendly affectionate relation may be said to "lower the threshold" for other responses such as sympathy. Or, conversely, as we have already suggested, the insecurity involved in lack of friends may raise the thresholds for sympathy which appears spontaneously when that insecurity vanishes in a new friendship.

An expression of identification of a less stable sort appears in the following episode:

3/15/34 Julius "shot" Saul with his toy gun.
Saul cried.
Julius said, "I shot him."

[1] Margaret Mead comments that among the cultures she has studied the Mndugomor are never sympathetic except when they are miserable or somewhat "taken down."

Julius ran off, then came back, shot again.
Saul cried.
Patrick watched, comforted Saul, and said, "Let's break his neck, too, Saul."
Saul said, "Let's break his neck, too."
Patrick said, "Then I'll kill him, too."
Saul said, "Yes, 'cause he's a naughty boy."
Both ride off talking.
Saul said, "Let's send him to the butcher store, O.K.?"

Here Patrick's response to Saul takes the form of a retributive fantasy in which he and the victim cooperate in retaliation upon the aggressor of his friend.

Other variations in behavior, judged from an adult point of view, were seen when the results of an experimental situation were compared with the day by day observations on the playground. A baby in a playpen, alone in the experimental room, served to redintegrate home patterns of protectiveness and play in Reinhardt who had made almost no contact with the crowd on the playground; while Janet, who scored highest on sympathetic and cooperative behavior on the playground, seemed to be completely at a loss in what was to her the queer situation of the experiment. She did little but stare at the child in the pen during the three-minute interval before the experimenter came in to give a suggestion. Contrast this with the following record of Julius, another child in Group H, whose playground score for sympathy was close to Janet's, and whose experimental score was also high.

First thirty seconds: Julius: "I'll come with you, shall I?" Starts to climb in, then: "I'll give you your doll, this is your dolly." Puts doll in pen.

Second thirty seconds: Julius: "And here's your teddy bear and here's your ball." Puts them in pen. Plays with auto.

Second minute: Julius: Puts auto in pen. "I'll get in with you." Climbs in. "Hello." Peers at Ann smiling. Laughs. Plays with auto. "This is where the gas goes, and this is where the water goes." Ann: "There." Lies down to look at auto. Julius plays with doll. "She opened her eyes."

The behavior of Janet and Julius in this experiment gave evidence for the experimenter's judgment that to Janet it was an unusual and adult-dominated situation which she did not understand while to Julius the situation consisted of a little girl in a playpen with all her toys outside; so he helped her first, then played with her as he would have done on the playground. This simply adds the final stone to the structure which has been built up to show that the interplay between external situation, the inner state of the child, and the symbolic values or meanings of the situation to him, rather than a mechanical carry-over of specific responses conditioned to isolated elements, determines behavior.

The differences in symbolic value of different situations to different children appeared dramatically in their responses to the experiments with pictures, and with animals. For instance, one picture of a child inside a low wire garden fence holding a rabbit was chosen as a fairly neutral picture but with the expectation that some children might think the rabbit was being injured or in danger of injury. Actually the gestalt apprehended by many children was quite different from that caught by the adults. The dark background recognized by adults as a photographic device was seen by some of the children as evidence that the child was *in* the dark. The fence had dire meaning for a little boy (Reinhardt) who was later discovered to have been left for hours at a time in a playpen.

Another picture showed a little barefooted boy crying, running out a gate, with children watching him, a dog in the foreground, and an automobile in the background. The responses to this were varied and suggested the indirect effect of projection through a selective influence on perception. The dog, the auto, the little girls, the boy's own bare feet stood out for different children in connection with the small boy's crying. Only one child appealed to an element not directly suggested by the picture—that is, the absent mother.

Certain interesting individual trends appeared in the responses to the pictures as a whole. Daniel, who was considered by teachers as the child who cried the most repeatedly, gave the crying child in the pictures which contained one, his own name. Others had certain persistent ideas—for instance, Kirk repeatedly gave "spanking" as the reason for a cry; Betsy repeatedly suggested fantastic reasons for a child's crying such as "The dog is going to bite the little boy's leg", "Maybe she broke her fingers."

In a series of experiments with animals, there were similar variations in situation-values. Few children had a generalized pattern of being kind to all animals, although most of the children had specific patterns of being "nice" to the rabbit or guinea pig. More strikingly, they responded in terms of values built up through previous experience; Wallis, just come from a farm, protested violently against hurting a chicken, where to less experienced children, the chicken was something to experiment with just as you explored and poked the turtle, or pulled a fish out of water to see what would happen. Here again, the newness of the situation, the presence of the adult, had different significance for different children, even though the experiments were carried out in a corner of the playground with as close a similarity to other nursery-school events as possible. Objective standardization of procedure gave little help toward subjective equalization of a situation approached by children so different in attitude. For each child the situation appeared to have a different value depending on his suggestibility or suspicion of adults, of the experimenter, his generalized fear of animals, or fear of the specific one present, the strength which this situation had, in competition with other possible activities on the playground, to hold his complete interest.

In sum, we see that in addition to general trends related to age and intelligence, specific characteristics of the situation determine behavior: the organization of the group, the relation between equipment and the number and interests of the children, specific dominant personalities among the children, and attitudes or behavior of the teachers. Variations in children's behavior and even apparent contradiction, if we look for statistical coherence, are to be understood in terms of the response of a given child at a particular time, in terms of his own psychological relation to the situation and the symbolic values of the specific situation to him at that time. These factors do not operate together in simple additive fashion but in complex ways illustrated by the fact that a motive or situation which increases the sympathy of some children prevents its appearance in others. From an educational point of view, in fact, it is at least as important to know what factors of interaction between the child and the situation are inhibiting or overstimulating sympathetic behavior— or any other aspect of social behavior—as it is to know what to expect of a three-year-old or a four-year-old of a given level of intelligence.

THE EDUCATIONAL ACHIEVEMENT OF PUPILS IN DIFFERENTIATED AND UNDIFFERENTIATED GROUPS*

A. G. BREIDENSTINE
Principal East Lampeter Township High School, Soudersburg, Penna.

INTRODUCTION

The purpose of this study is to present evidence concerning the validity of a modern educational practice in certain schools, designed to improve the educational achievement of the pupils in those schools. This practice, whose worth is conjectural, is that of the grouping of children in a school system on the basis of ability. Shall children be grouped according to some particular scheme, or shall they be taught in groups of reasonable size without any attention to the membership of the group? Grouping by grade has been in practice for many years and will not, therefore, be challenged here. More recently, however, finer differentiations have been attempted; some groupings are based on such measures as intelligence quotients and teachers' marks; others, on a certain average or combination of measures.

These various devices are being used in many ways and for different purposes. A rather general reason for their use has been to improve the educational achievement of the pupils concerned. Do pupils grouped according to the devices now employed show greater progress in their educational achievement than those taught in the traditional way, if the factors of intelligence and grade-placement are balanced for both groups? This problem falls within the scope of the present survey. In brief, the study seeks to present evidence concerning the educational achievement of both the differentiated and the undifferentiated groups in grades two to nine inclusive.

I. THE STATUS OF DIFFERENTIATED GROUPING

A. Definition of Differentiated Grouping

Differentiation is the grouping of pupils for the purpose of effecting greater homogeneity. Any plan of differentiation must consider the two opposite poles of the situation. At the one pole is complete heterogeneity, with no attempt whatever at differentiation; at the opposite pole is complete homogeneity, or perfect differentiation. The former plan would group all children with no attempt to section according to chronological age, mental age, previous achievement, or any other factor. It is evident in the light of present investigations and procedures that no differentiation at all would be undesirable and, on the other hand, that differentiation in order to be perfect would have to take into account every factor affecting the learning situation so as to equalize perfectly all members of any one group. The latter method is impossible in any group save that group defined by Turney as a "theoretically homogeneous 'group' of one".[1] Most educators realize that such grouping would lead to individual instruction and hence would deny the postulate that differentiation implies group instruction. Beliefs in differentiation and undifferentiation are, therefore, relative; that is, they do not support wholly either extreme. To what degree we might wish to retain heterogeneity will depend largely on our objectives for either plan.

F. J. Barnard said in 1899: "The basic idea in the education of our youth is mental activity. We need therefore to classify pupils so as to shorten the way for the quick-

* Doctor's dissertation, Temple University, May, 1936. Thanks are especially due the principals whose cooperation in their individual schools made available the data utilized in this study. Acknowledgment is made of technical advice and criticisms received from George E. Walk, Frederick H. Lund, I. S. Hofier, Ralph D. Owen, R. W. Schlosser, and Rebekah S. Sheaffer.

[1] Austin H. Turney, "The Status of Ability Grouping," *Educational Administration and Supervision*, XVII (January, February, 1931), 21–42, 110–27.

witted and permit them to enter new fields. The slow thinkers need more attention. Restrain neither but enable both to make progress at different rates of speed."[2] The primary objective in 1899 was that of completing a course in record time so as to permit employment, or further study, or any other activity. Whatever the secondary objectives may be to-day, educators generally agree that the objective of educational achievement is primary. Symonds says that the aim of ability grouping is, "the bringing together of pupils who will be alike in achievement at the end of the period of learning."[3] Such an aim approximates the objective of to-day which is none other than greater educational achievement. Practically all of the experiments which will be referred to later in this study hold the latter objective as the more important.

B. *Alleged Advantages and Disadvantages*

Austin H. Turney, who has summarized many of the studies on grouping in *Educational Administration and Supervision* (January and February, 1931), lists the alleged advantages and disadvantages as proposed by Ashbaugh[4], Briggs[5], Clerk[6], Freeman[7], Miller[8], and Seashore[9].

The advantages claimed for differentiated grouping are as follows:

1. It permits pupils to make progress commensurate with their abilities.
2. It makes possible an adaptation of the technique of instruction to the needs of the group.
3. It reduces failures.
4. It reduces elimination.
5. It helps to maintain interest and incentive, because bright students are not bored by the participation of the dull.
6. Slower pupils participate more when not eclipsed by those much brighter. The confidence of the slow pupils is restored.
7. It may make competition operative as an incentive.
8. It makes teaching easier.
9. It makes possible individual instruction to small slow groups.

The disadvantages ascribed to differentiated grouping are that:

1. Slow pupils need the presence of the able students to stimulate them and encourage them.
2. Ambitious pupils overwork.
3. Bright pupils graduate too young.
4. A stigma is attached to low sections, operating to discourage the pupils in these sections.
5. Teachers are unable, or do not have time, to differentiate the work for different levels of ability.
6. Teachers object to the slower groups.
7. Parents complain when their children are placed in slower sections.
8. Program construction is rendered more difficult.
9. Frequent transfers necessitate more efficient office help.

C. *Bases for Differentiation*

In the present study on grouping, as in many earlier studies, the bases for grouping are numerous. Some of the administrators cooperating in this study used as a basis for sectioning only the intelligence quotient of the students concerned; others employed past achievement plus the intelligence quotient: while still others used the foregoing criteria and the subjective measure of a teacher's recommendation. Kefauver says, "Whatever the combination of factors used for distributing pupils to ability groups it should contain either a composite of the marks in the elementary school or a rating of capacity by the teacher."[10] Brooks concluded after some ex-

[2] F. J. Barnard, "Let Pupils Be So Classified as to Allow Unrestricted Progress or Unlimited Time According to Ability." *Proceedings National Educational Association*, XXXVIII (1899), 163–70.
[3] P. M. Symonds, *Measurement in Secondary Education*. New York: The MacMillan Company, 1928.
[4] E. J. Ashbaugh, "Homogeneous or Non-Homogeneous Grouping," *Journal of Educational Research*, IX (March, 1924), 241–45.
[5] T. H. Briggs, "Provisions for Abilities by Means of Homogeneous Groupings," *Third Yearbook of the National Association of Secondary School Principals*. Menasha, Wis.: George Banta Publishing Co. Pp. xxxi + 87.
[6] F. E. Clerk, "The Arlington Plan of Grouping Pupils According to Ability in the Arlington High School, Arlington, Mass.." *School Review*, XXV (January, 1917), 26–47.
[7] F. N. Freeman, "Sorting the Students," *Educational Review*, LXVIII (November, 1924), 169–74.
[8] W. S. Miller, "General Intelligence Tests," *School Review*, XXVIII (February, 1920), 94.
W. S. Miller, "The Administrative Uses of Intelligence Tests in the High School," *Twenty-first Yearbook of the National Society for the Study of Education*, Part II. Bloomington, Illinois: Public School Publishing Co., 1922.
[9] C. E. Seashore, "Sectioning Classes on the Basis of Ability." *School and Society*, XV (1922), 353–58.
[10] Grayson N. Kefauver, "The Validity of Bases for Forming Ability Groups," *Teachers College Record*, XXXI (November, 1929), 99–115.

perimenting that sixth grade marks are as good an index of seventh grade accomplishment as any one of seven mental tests, or better than any one, with the exception of the Stanford Revision of the Binet test.[11] Since there is usually a high correlation between a pupil's achievement and his mental ability, this proves very little because the measure of ability was really included in the achievement test, although obscured. The Binet test, however, which is usually conceded to measure untaught facts only, would point to exceptions and hence would not indicate exactly the same results.

The researches of Kelley[12], Pressey[13], Stone[14], Hughes[15], Poffenberger and Carpenter[16], Fleming[17], Pierson and Nettels[18], and Hughes[19] all refer to factors other than intelligence which can be used in differentiation. Billett[20], however, concludes that of all the bases yet used the best measure of mental ability seems to be the best single basis. Ryan and Crecelius[21] have added many new bases to-gether with the common ones: chronological age, mental age, intelligence quotient, pedagogical age, social age, height, weight, general conditions of health, and rank in class. By using all these bases we come rather close to the "theoretically homogeneous 'group' of one."

D. *Criteria of Success of Differentiated Grouping*

Austin H. Turney lists the criteria as he found them in published reports.[22]

1. The comparative achievement of pupils of equal ability in homogeneous and heterogeneous sections.
2. The effect upon failure and elimination.
3. The mental hygiene or happiness of the pupil.
4. The motivation of the pupil.
5. The ease of teaching and teachers' attitudes toward ability grouping.
6. The comparative achievement of classes of different ability.
7. The amount of shifting following ability grouping.

In the study undertaken here only the first criterion has been considered. The others listed above would provide fruitful research for those interested in this subject. In summarizing the researches previously made, most attention will be given to this criterion, although occasionally the others will obtain.

E. *Summaries of Researches Previously Made*

Numerous researches have repeatedly attempted to solve the problems relating to differentiated grouping. Those to be summarized here are concerned especially with the problem of educational achievement. The experiments involving the other criteria previously mentioned will be listed in the bibliography, but, inasmuch as they are not directly related to the present investigation, no summary of them will be given here.

Barthelmess and Boyer conducted an experiment in the Philadelphia elementary schools to determine the value of ability grouping as it pertains to educational achievement. The conclusions from this carefully controlled experiment are as follows:

1. There is a statistically significant difference in favor of homogeneously grouped pupils as far as the improvement of arithme-

---

[11] F. D. Brooks, "Sectioning Junior High School Pupils by Tests and School Marks," *Journal of Educational Research*, XII (December, 1925), 350–69.

[12] T. L. Kelley, *Educational Guidance*. Teachers College Contributions to Education, No. 71. New York: Teachers College, Columbia University, 1912.

[13] S. L. Pressey, "An Attempt to Measure the Comparative Importance of General Intelligence and General Character Traits in Contributing to Success in School," *Elementary School Journal* (November, 1920), 220–27.

[14] C. L. Stone, "Disparity Between Intelligence and Scholarship," *Journal of Educational Psychology*, XIII (April, 1922), 241–44.

[15] W. Hardin Hughes, "Organized Personnel Research and Its Bearing on High School Problems," *Journal of Educational Research*, X (December, 1924), 386–98.

[16] A. L. Poffenberger and F. L. Carpenter, "Character Traits in School Success," *Journal of Experimental Psychology*, VII (February, 1924), 67–74.

[17] C. W. Fleming, *A Detailed Analysis of Achievement in the High School*. Teachers College Contributions to Education, No. 196. New York: Teachers College, Columbia University, 1925. Pp. xi + 209.

[18] C. D. Pierson and C. H. Nettels, "A Study of High School Seniors to Determine Who Shall be Recommended to College," *School and Society*. XXVIII (August, 1928), 215–216.

[19] J. M. Hughes, "A Study of Intelligence and of the Training of Teachers as Factors Conditioning the Achievement of Pupils. II." *School Review*, XXXIII (April, 1925), 292–302.

[20] R. O. Billett, "A Controlled Experiment to Determine the Advantages of Homogeneous Grouping," *Educational Research Bulletin*, (April 4, 1928–May 2, 1928).

[21] H. H. Ryan and P. Crecelius, *Ability Grouping in the Junior High School*. New York: Harcourt Brace, 1927. Pp. 223.

[22] Austin H. Turney, "The Status of Ability Grouping," *Educational Administration and Supervision*, XVII (January, February, 1931), 21–42, 110–27.

tic, reading, and technical English skills is concerned.

2. Since ability grouping is only one of the many factors which facilitate improvement, a very small amount of superiority may be very significant.

3. Improvement was found to exist in each of the several groups, the high, the low, and the medium.[23]

R. O. Billett performed an experiment extending over a period of three years in the Harvey High School of Painesville, Ohio. His subjects were three successive groups of ninth-grade pupils in English. The first year the pupils were sectioned on the basis of intelligence quotients derived from the Terman Group Test of Mental Ability, and for the second year on the basis of combined scores of mental ability and achievement. The third year they were again sectioned on a basis similar to the method used in the first year. The groups included in the experiment were a fast homogeneous group, an average homogeneous group, and a slow homogeneous group with two heterogeneous groups as controls. The following factors were controlled in the experiment: (1) course of study, (2) time allowed, (3) teacher (same for all), and (4) measuring instruments.

The conclusions are as follows:

1. Teachers' ratings or teachers' marks should never be used in toto or in part as a basis for homogeneous grouping, because such ratings or marks are largely measures of qualities which education seeks to change.

2. Homogeneous grouping for the purposes of education should be based on traits not subject to change through education, such as, the intelligence quotient, the probable learning rate, or the index of brightness.

3. Grouping on the basis of the intelligence quotient has shown a marked measurable advantage for slow pupils, some measurable advantage for the average pupils, and slight measurable disadvantage for the bright pupils.

4. The slight measurable loss to bright pupils because of homogeneous grouping is far outweighed by the measurable advantages which seem to affect the records of the slow and average pupils.[24]

H. S. Bonar conducted an experiment at Manitowoc, Wisconsin, to prove, if possible, the value of differentiated grouping. The experiment was begun in 1927 in three elementary schools. The purposes and results of one year of the experiment follow. The experiment was conducted to determine the merits of differentiated grouping so as better to meet the objections of teachers who generally prefer not to teach the slow groups, and to answer the criticisms of parents who contend that the practice is not democratic. Classes were organized six weeks after school had opened so as to include teachers' judgments in the classification. The Detroit First-Grade Intelligence Quotient Test, the Herring Revision of the Binet–Simon Test, the judgment of the kindergarten teacher, and the judgment of the first grade teacher after the six weeks of school were collectively used in the differentiation.

Class I, of thirty-six children in one school, was the highest section. Class II was the section containing all abilities. Class III was the low section. Classes II and III had mean intelligence quotients approximately equal, Class II having 99 and Class III, 97.5. On the Gates Silent Reading Test the accomplishment in Class III was much lower than the difference in the mean intelligence quotients would warrant. To be sure, different and varying factors enter into the results of any experimental adventure.

The teachers constitute one such variable. However, there were employed in the Bonar research only seasoned teachers using the same materials for motivation and the same modes of encouragement under the same supervision, and having the same training and experience. The significant findings are outlined herewith:

1. No advantage accrues from differentiation either in the accelerated or the retarded groups.

2. There was sufficient evidence to show that the greatest accomplishment takes place in mixed groups, ability considered.

3. The classification of children in the lower grades so that all abilities are represented eliminates the problem of teachers who

[23] Harriet M. Barthelmess and Philip A. Boyer, "An Evaluation of Ability Grouping," *Journal of Educational Research*, XXVI (1932), 284–94.

[24] R. O. Billett, "A Controlled Experiment to Determine the Advantages of Homogeneous Grouping," *Educational Research Bulletin*, VII (April 4–May 2, 1928).

object to teaching only slow pupils and prevents the criticism that differentiated grouping is undemocratic.[25]

H. E. Burt, L. M. Chassell, and E. M. Hatch conducted an experiment in differentiation by sectioning college freshmen in psychology. The sections were formed on the basis of the Army Alpha Intelligence Test results. Three groups were formed, and for every selected group there was at least one control group. Three instructors participated in the experiment, keeping the content and teaching similar for all students included in the study.

During the first semester the rate of progress was practically the same for all sections. During the second semester the high group, and the medium group, with their two controls, were pushed. The high group covered the work in twenty-seven instead of the usual thirty-seven hours. The medium group covered the course in thirty hours instead of the usual thirty-seven. The low section was not pushed. The results of the experiment were studied by means of a correlation technique. The persons investigating this earlier study worked according to the theory that if instruction is such that all students are urged to do their utmost, then the correlation between intelligence and academic marks will be high. This is exactly what they found in the experiment. In addition, the saving of time, previously mentioned, was notably a direct outcome of differentiation, according to the Burt, Chassell, and Hatch method.[26]

R. R. Cook conducted an experiment in differentiation involving high-school classes in English I, English III, plane geometry, and ancient history. Pupils were sectioned on the bases of intelligence tests and marks made in the preceding semester. The teachers who participated in the study were interested, but made no attempt to adapt special teaching methods. The results were as follows:

1. Pupils of superior ability in geometry did better work in mixed classes than in selected groups.

2. The ancient history scores of the pupils in selected groups were much higher than the scores of pupils having the same ability in mixed classes.

3. In English the differences were negligible.

4. The poor pupils did better in geometry, slightly better in English, but in history not nearly so well when in separate groups.

Cook concludes by saying, "We are unable to say that the grouping of high-school children according to ability secures better results than random grouping".[27]

August Dvorak and J. J. Rae report an experiment in which they tested first-grade children by means of three tests, the Pintner–Cunningham Primary Mental Test, the Holley Picture Completion Test for Primary Grades, and the Detroit First-Grade Intelligence Test. On the bases of these tests three groups were formed, two equivalent groups and one segregated group. The following observations obtained:

1. When methods and materials of instruction are adapted to abilities, results are positive.

2. Mere segregation of bright pupils into homogeneous groups without adaptation produces negative or indifferent results.

3. We need to discover techniques and materials for various abilities of pupils in homogeneous groups.[28]

Anna M. Engel compared the high-school records of 139 graduates from special advanced classes in the Detroit schools with 127 control pupils who had not entered the special classes. The results as given are not very significant. She concludes that:

1. The differences in algebra and Latin were inconclusive.

2. A noticeable superiority was found on the part of the special-class graduates in English.[29]

William A. McCall conducted an investigation in which sixty-seven bright children who were taught in segregated classes were matched with sixty-seven bright children from unsegregated classes. The mental and the

[25] H. S. Bonar, "Ability Grouping in the First Grade," *Elementary School Journal*, XXIX (May, 1929), 703–6.

[26] H. E. Burtt, L. M. Chassell, and E. M. Hatch, "Efficiency of Instruction in Unselected Sections in Elementary Psychology Compared with That in Sections Selected on Basis of Intelligence Tests," *Journal of Educational Psychology*, XIV (March, 1923), 154–61.

[27] R. R. Cook, "A Study of the Results of Homogeneous Grouping of Abilities in High-School Classes," *Twenty-third Yearbook of the National Society for the Study of Education*. Bloomington, Illinois: Public School Publishing Co., 1924.

[28] August Dvorak and J. J. Rae, "A Comparison of the Achievements of Superior Children in Segregated and Unsegregated First-Grade Classes," *Elementary School Journal*, XXIX (January, 1929), 380–86.

[29] Anna M. Engel, "Comparison of Class Ratings of Pupils in Special Advanced Classes with Accelerated Pupils in Regular Classes in the Detroit Public Schools," *Twenty-third Yearbook of the National Society for the Study of Education, Part I*, pp. 297–301. Bloomington, Illinois: Public School Publishing Co., 1924.

chronological ages of every pupil in one group matched the mental and the chronological ages of a corresponding individual in the second group. Likewise, group one as a whole matched group two. The tests used to measure the educational achievement of the sectioned classes were the Thorndike–McCall Reading Scale, the Woody–McCall Mixed Fundamentals in Arithmetic, and selected words from the Ayres Spelling Scale. All of these tests were given at the end of two years of segregation. The results are as follows:

1. The segregated group had gained an excess of 7.14 months in the two years. We can be sixty-two per cent certain, therefore, that such segregation is favorable to the growth of the traits measured.[30]

Edward L. Moyer compared the educational achievement of high-school pupils who were grouped according to ability with those who were in mixed classes. The achievement was measured by standardized tests and teachers' marks. When the comparisons of achievement were analyzed the following results were obtained:

1. In algebra there was no difference in achievement.
2. In Latin the bright pupils in superior segregated classes excelled the bright pupils of mixed classes.
3. In both subjects the medium pupils in superior or medium segregated classes excelled the medium pupils of mixed classes.
4. If pupils are grouped by ability, those who are misplaced in higher sections profit and those who are misplaced in lower sections suffer.[31]

T. L. Purdom studied the achievement of high school pupils in algebra and English. The achievement of pupils in differentiated groups was compared with the achievement of pupils in undifferentiated groups. The pupils in both groups were matched for age, sex, and intelligence score. The experimental group and the control group had the same teacher throughout the experiment. Purdom's results show very little in favor of any group. When the grades of the five participating schools were treated as a unit, no pronounced differences were evident. When treated singly, by subject, the semester grades of the differentiated sections were lower in English but higher in algebra than the corresponding grades of undifferentiated pupils.[32]

J. M. Shields performed an experiment on the teaching of reading in differentiated groups. The bases for segregation were the speed of reading and the comprehension of reading. Four groups were formed corresponding to the four quadrants of a scattergraph in which the ordinates represented speed and the abscissae comprehension. The four groups were as follows: high in both speed and comprehension, high in comprehension and low in speed, high in speed and low in comprehension, and low in both speed and comprehension.

The pupils in these groups were then paired with pupils in undifferentiated groups, matching them in chronological age, mental age, and intelligence quotient. A strong teacher was placed in charge of the controls while an experimenter taught the experimental classes. At the end of six weeks the tests given at first were given again. Only twenty-nine pairs remained through the experiment. The results were decidedly in favor of the experimental pupils who were above and below the average in intelligence quotient and in chronological age. There was little difference between those of average intelligence quotient and chronological age. Quantitatively the results were:

1. The experimentals gained 42.6 points in speed.
2. The controls gained 27.2 points in speed.
3. The experimentals gained 3.4 points in comprehension.
4. The controls gained 1.4 points in comprehension.[33]

M. J. Van Wagenen formed three types of groups for his experiment on achievement quotients of pupils. His three groups were first, a differentiated class of superior chil-

[30] Wm. A. McCall, "A Comparison of the Educational Progress of Bright Pupils in Accelerated and in Regular Classes." *Twenty-Seventh Yearbook of the National Society for the Study of Education*, Part II. Bloomington, Illinois: Public School Publishing Co., 1928.

[31] Edward L. Moyer, "A Study of the Effects of Classification by Intelligence Tests," *Twenty-Third Yearbook of the National Society for the Study of Education*. Bloomington, Illinois: Public School Publishing Co., 1924.

[32] T. L. Purdom, *The Value of Homogeneous Grouping*. Baltimore: Warwick and York, 1929. Pp. 96.

[33] J. M. Shields, "Teaching Reading through Ability Grouping," *Journal of Educational Method*, VII (September, 1927), 7–10.

dren; second, a mixed class of superior and non-superior children in about equal numbers; third, a composite class of regular pupils. The achievement quotients were compared in all cases and the results were thus recorded by Van Wagenen: "In no phase of instruction are the differences marked enough and consistent enough to suggest any distinct advantage for any one of the three types of classes under the conditions which prevailed in this school system."[34]

The experiment of J. T. Worlton included 3700 pupils distributed through grades IVB, VB, VIB, and VIIB in twenty-four schools. Sixty-seven teachers taught 2052 pupils in homogeneous classes, the remainder being taught in control groups by fifty-two teachers in heterogeneous classes. The conclusions of the experiment embrace three noteworthy observations:

1. In the grammar grades those scoring ninety and above in native ability profited noticeably by classification.

2. Homogeneous classification of pupils offered advantages over heterogeneous classification in the scholastic achievement of pupils of normal ability and more than normal ability.

3. Measures showed that pupils are now more than ever before achieving scores commensurate with their abilities.[35]

II. The Nature of the Present Investigation

A. *Statement of the Problem*

The educational achievement of pupils in the public schools reveals the problem of individual differences. Educators have been alert to this problem for many years. Some have been led to the conclusion that the only solution is that of individualized instruction; others seek a solution in bringing together into groups of fair homogeneity those pupils who have similar learning characteristics. Such grouping, it is claimed by a certain school of modern educators, would improve educational achievement and at the same time result in other desirable outcomes. Others strongly deny that any values whatsoever would accrue from selective grouping. It is the aim of this investigation to throw some light upon the controversy so that school administrators may have statistical evidence either for or against the practice of grouping. To arrive at a valid and reliable criterion of educational achievement, tests

Tabulation of Studies Reporting Gains, Losses, and Inconclusive Results in the Case of Segregated Pupils as Compared with Unsegregated Pupils[36]

Levels of Ability	Gain	Loss	Inconclusive
Inferior Pupils	1. Billett, English 2. Cook, geometry 3. Cook, English II and III 4. Shields, English	1. Cook, ancient history	1. Bonar, first grade reading 2. Moyer, algebra
Superior Pupils	5. Shields, English 6. Bonar, reading (slight) 7. Cook, ancient history 8. Engel, English 9. McCall, educational age 10. Worlton, three subjects 11. Moyer, Latin	2. Billett (slight) English 3. Cook, geometry	3. Bonar, first grade reading 4. Cook, English 5. Engel, algebra 6. Engel, Latin 7. Van Wagenen, Achievement Quotient, various subjects 8. Moyer, algebra
Average Pupils	12. Billett, English 13. Worlton, three subjects 14. Moyer, algebra		9. Shields, English
All Levels	15. Purdom, algebra (slight) 16. Barthelmess and Boyer (skills)	4. Purdom, English	10. Purdom, test scores, for algebra and Latin

[36] This tabulation, with the exception of the finding by Barthelmess and Boyer, is the conclusion of Austin H. Turney as found in *Educational Administration and Supervision* for January and February, 1931.

[34] M. J. Van Wagenen, "Effect of Homogeneous Grouping upon the Quality of Work of Superior Children," *Journal of Educational Method*, VI (February, 1927), 240–47.

[35] J. T. Worlton, "The Effect of Homogeneous Classification on Scholastic Achievement of Bright Pupils," *Elementary School Journal*, XXVIII (January, 1928), 336–45.
J. T. Worlton, "Why of Homogeneous Classification," *Elementary School Journal*, XXVII (December, 1926), 265–74.

were applied to all levels of intelligence over a reasonably large cross-section of our school system. These tests furnished data which, when treated by means of statistical devices (to be explained later), served as the basis of certain conclusions. Grades were studied individually and collectively so as to detect any possible justification for grouping or for the retaining of heterogeneous classes. Educational achievement was never studied by itself, but always in relation to the intelligence behind it. Is the educational achievement among pupils in differentiated groups greater than that of pupils of undifferentiated groups, provided that the individuals of both groups are matched in intelligence and grade-placement in school? This is the problem to be solved.

B. *Delimitation of the Field of Investigation*

Whatever the outcomes of differentiated grouping may be, this investigation is concerned only with educational achievement. The achievement of pupils in only grades two to nine inclusive was studied, thus necessitating only one type of test and making it possible for all scores to be transmuted into values possessing equal significance.

The subjects for this investigation were selected from eleven schools which indicated their willingness to participate. No attempt was made to select schools for any reason other than experience in the practice of grouping and an enthusiastic support on the part of the principal so as to help supervise and administer the details of the testing program involved. The problem in its limited form is, therefore, essentially the same as the larger problem of the study. The former differs from the latter only in relation to grades studied, the number of schools, and the number of school subjects included in the investigation.

C. *Timeliness of the Problem*

Educators have long been concerned with the number of repetitions necessary for bright and dull pupils. Those who advocate segregation claim that the bright pupils are annoyed by the repetitions necessary for permanent learning by the dull. Authorities agree, however, that repetitions are essential in order to spare the retarded child the embarrassment of confusion and negativism. Those opposing segregation claim that the dull child benefits in his contacts with those of higher intelligence. Teachers, it is claimed, in turn do their best work in heterogeneous classes. Thus the contention of the latter group is that the benefits of undifferentiated grouping accrue to both the bright and slow pupils.

Teachers, supervisors, and principals are always seeking methods of instruction which prove more efficient than those now in practice. Many ask whether or not differentiation would provide a situation making possible the adaptation of specially designed techniques of teaching. School officials, patrons, and taxpayers are often concerned with the costs attending failure. "Would differentiation reduce the number of failures?" is their pertinent question. Some blame the school for losing many of the pupils who, because they are poorly adjusted to the curriculum requirements, lose interest and are consequently removed from school. Would grouping aid in the solution of this problem? All of us are aware of increased enrollments, new compulsory attendance laws, recent child-labor laws, and hence the great number of pupils who, against their will, are deposited in our schools. All of these factors are related to the achievement of the pupil in school and are therefore germane to the subject of grouping.

III. Procedure and Development

A. *Original Contacts*

In order that any study involving a testing program may be successful, whole-hearted cooperation in every detail is necessary. Realizing this, the writer contacted school administrators who were willing to participate in such a study as is herein described. Only those who declared themselves enthusiastic were selected, all of whom responded in a fashion quite unusual for busy school officials. Schools were selected first, therefore, according to the willingness of the administrators to cooperate; second, according to the grades included in the organizations; and third, according to the practice of the school relative to grouping, whether differentiated or undifferentiated. Eleven schools participated in the study, four practicing differentiation and seven undifferentiation. These schools provided for the study 2,325 children from grades two to nine inclusive. Of the original complete returns, 1,163 pupils were taught in undifferentiated classes while 1,162 were taught in differentiated classes. In every one of the schools cooperating, the principal per-

sonally supervised for the study every phase of the testing program. Great precaution was taken in the administering of the tests and in no case was an inexperienced person permitted to handle the work. In every school either the supervising principal or his assistant administered the tests according to the directions given in the test instructions. So profoundly interested were the cooperating principals that six not only gave the tests, as they had agreed to do, but also scored them with a view to keeping the scores for their own use in their school records. Those who did not score the tests have since either secured the scores or requested their remittance. In every detail, therefore, the data used in this study served a useful purpose not only in this treatise on differentiation, but also in the records and guidance departments of the various schools concerned. The cooperation of the several school officials and their constant interest in the prosecution of the present survey contributed in a large way to the thoroughness with which this first step of the statistical study on differentiation was concluded.

The Selection of the Tests

After the arrangements with the eleven schools had been completed, it was necessary, as a second step, to select and distribute the tests which were to be used. These tests were of two kinds: first, the Group Intelligence tests; and, second, the Educational Survey tests. Only those schools which had not previously determined the intelligence quotients by means of some reliable test were asked to administer the intelligence tests. The Otis Group Intelligence Scales, Primary Form A, were used for the lower grades while the Otis Self-Administering Tests of Mental Ability, Intermediate Form A and Higher Form A, were used for the higher grades in which the tests were suited to the levels of the respective groups. For every pupil, therefore, the most recently determined intelligence test results were employed. For the educational achievement survey the New Stanford Achievement Tests were used. The Primary Examination, consisting of the tests on Paragraph Meaning, Word Meaning, Dictation, Arithmetic Reasoning, and Arithmetic Computation, was employed for grades two and three. The Advanced Examination, consisting of the tests on Paragraph Meaning, Word Meaning, Dictation, Language Usage, Literature, History and Civics, Geography, Physiology and Hygiene, Arithmetic Reasoning, and Arithmetic Computation, was used for grades four to nine inclusive. The New Stanford Achievement Test was selected because of its comprehensive nature and its provision for various other possible interpretations of the test score, such as, educational age, educational quotient, and accomplishment ratio (which is determined by dividing the educational quotient by the intelligence quotient). As a secondary consideration, nevertheless an important one, these tests were selected because none of the eleven schools had administered them before to any one of the pupils included in the study. In only one school had they ever been used at all. As for the reliability of the tests, the authors, Truman L. Kelley, Giles M. Ruch, and Lewis M. Terman, conclude from their experience with the tests that if the results of their three-hour examination are combined into an all-round educational age, it is entirely justifiable to place a great deal of confidence in the measure. Composite scores should, they contend, be used in all surveys. The present survey used composite scores.

Distribution and Administration of Tests

The tests were sent to the participating schools so that the tests might be administered during the month of March, 1935. All the schools concentrated their testing into a period of a little more than three weeks. This factor is not of great importance since the age of the pupil and the total score were always used together throughout the study. However, it is a double precaution to take care of the effect of such educational growth as might result from special application on the part of the student toward the end of the school year, which might have some significance.

The correct administering of the tests was vouched for by every school official cooperating in the study. For practically all the second-grade pupils, the teachers filled in the pupil's name, the grade, the sex, the age, and the date of taking the test so as to avoid errors on the part of the immature pupil in taking directions. Those giving the primary tests usually divided the test period into two sittings to avoid excessive fatigue. The advanced tests were usually given at four sittings. During the giving of the tests in the larger schools two teachers, the principal and

the classroom teacher, jointly administered the tests. After the tests had been administered they were returned to the office where the principal or his helpers recorded the intelligence quotients of the pupils on the achievement test blanks. They were then returned to the writer for scoring or, as in several instances referred to, first scored and then returned after the clerks had recorded the pupil's score and his educational age. About one-third of the test scores were checked to insure accuracy, although the few errors discovered usually proved negligible when the individual test scores were resolved into the composite examination score. Scoring errors and errors due to faulty administration were, therefore, reduced to a negligible minimum throughout.

Devices and Methods of Measurement

After all the tests had been corrected and checked, the following items were recorded on small cards to facilitate handling: name, sex, intelligence quotient, composite score, grade in school, educational age, actual age, educational quotient, accomplishment ratio, type of grouping, and name of school. The name, sex, intelligence quotient, actual age, and name of the school were merely recorded from the test blank. The reader will recall that the school officials recorded the intelligence quotient for every child on the achievement test blank. The actual age, however, was converted from years and months to years and the decimal part of a year correct to two decimal places. The educational age read from the test blank was dependent upon the achievement score. This age like the actual age was also changed to years and the decimal fraction of a year. The reason for such a procedure is obvious when it is realized that the educational quotient is the quotient obtained when the educational age is divided by the actual age. The computation involved in finding the educational quotient is similar to that of finding the intelligence quotient in which all decimals are dropped. Hence the educational quotient is always expressed in a whole number. It is a value quite equal to the intelligence quotient numerically. The accomplishment ratio is the measure designed to discover whether or not the achievement of the pupil is commensurate with his actual intelligence. The accomplishment ratio is found by dividing the educational quotient by the intelligence quotient. All of these necessary figures were then placed on regulation cards together with the type of grouping which was ascertained personally from the cooperating principals. From this stage on, only the cards were used except in cases of checking, or for obvious omissions. These cards were then sorted according to the grade of the pupil listed and the type of the grouping represented.

Having completed the recording and the sorting of the pupils, the investigator was then able to match the undifferentiated pupils with the differentiated in grade and in intelligence. Both the undifferentiated and the differentiated pupils' intelligence quotients were tabulated in frequency series using the same columnar interval for each particular grade. Upon comparison it was evident at a glance which interval of the groups had the greater frequency. Cards from both series were then eliminated until every interval of the one series had exactly the same frequency as that of the other. Prior to this process of matching, the median intelligence quotient for every group in every grade was determined. After the matching, the median intelligence quotient, to be sure, was the same for both groups. The former median was calculated only to make the comparison, and to determine roughly the original status of the group. It should be noted from now on in this discussion that whenever reference is made to the two groups, the differentiated and the undifferentiated, the writer refers to those groups which have previously been matched in intelligence and in grade.

The measures applied to describe the accomplishment ratios of the two groups were the median, the first-quartile, the third-quartile, the mean, the mode, and the standard deviation (as a measure of variability). While it is true that some of these measures had little signficance, they served, nevertheless, as checks on the more significant measures and helped in the completion of the entire statistical picture.

The relationship existing between the educational quotient and the intelligence quotient was studied by a correlation technique. Before using the correlation technique, however, the test of linearity was applied in each grade. The point of straight line relationship having been established, the correlation between the intelligence quotient and the educational quo-

tient was calculated by means of the Pearson product-moment formula. To present this relationship graphically both regression lines were superimposed on the scatter-diagram (see Appendix).

The measures applied, therefore, to study the educational achievement of the two groups concerned were those of central tendency, variability, and relationship.

Second-Grade Findings

Originally there were eighty-nine undifferentiated and sixty-four differentiated pupils of the second grade included in this study. Matching these in intelligence reduced the number in each group to sixty-one. These groups, matched for grade and intelligence, formed the bases for all the data of the second grade.

These second-grade groups presented only a reasonably symmetrical distribution of traits. The skewness of their distributions for the accomplishment ratios was found positive and according to Pearson's measure of skewness $S_k = \frac{M - M_0}{S.D.}$ equalled .3 for the undifferentiated and .1 for the differentiated.[37] Their median intelligence quotient of 107, however, was fairly high, especially in view of the fact that up to the time of the study no elimination due to incapacity to do school work could have taken place. Examining the two groups from the standpoint of dispersion, one observes significantly that for both the accomplishment-ratio distribution and the composite-score distribution the standard deviations were approximately equal in each case. The fact having been established that both groups were quite equally paired, the results obtained should, therefore, accurately describe the educational status of these groups.

The measures used to describe this educational achievement were the accomplishment ratio, the educational quotient, and the composite score. In order to facilitate comparisons the accomplishment ratio was converted into an accomplishment quotient by moving the decimal point as in the cases of the educational quotient and the intelligence quotient. To clarify these measures and their relative importance the following example may prove helpful. A child ten years old mentally and chronologically will have an intelligence quotient of one hundred. In a similar sense, if this child is also ten years educationally, he will have an educational quotient of one hundred. If then the educational quotient of one hundred is divided by the intelligence quotient (also one hundred), the ratio will naturally be unity. Instead, however, this ratio too was considered as the equivalent of one hundred for the purpose of facilitating the interpretation of all values.

The accomplishment of the two matched groups was determined for the first-quartile, for the median, and for the third-quartile.[38] This procedure was followed so that the various levels of achievement could be compared, as well as the average levels. For the groups as a whole, however, the arithmetic mean of the accomplishment ratio was used. The results very clearly indicate that the undifferentiated group excelled at every level and hence in the average, the difference of the mean accomplishment ratios of the two groups being 8.2 ± 2.0. Such difference did not arise from chance but as a result of factors which made the undifferentiated group decidedly superior.[39] Whether or not differentiation is the important factor in this variation will be more definitely shown when all the other grades have been similarly studied.

The educational quotient was used in like manner, with only one exception. Instead of using both the mean and the median to describe the central tendency, only the mean was used. As in the results obtained in the study of the accomplishment ratios, the undifferentiated group was again superior in the educational quotients. The difference of the means was 9.5 ± 1.8, a very significant difference in favor of the undifferentiated group.

Still another comparison, using only the mean composite scores, was made. It must be remembered that by the composite score is meant the average of all the individual subject scores. The difference between the means in this case was found to be $7.1 \pm .9$, establishing once more the superiority of the undifferentiated group.

---

[37] The skewness, the mode, and the standard deviation were found throughout in order that the statistical description might be more complete, and that those wishing to check the accuracy of this research might employ these measures accordingly.

[38] The first-quartile score in the distribution is that score which exceeds twenty-five percent of the scores; the third-quartile score exceeds seventy-five percent of the scores.

[39] When the difference is greater than four times its probable error, the difference is statistically significant. This test of significance obtains throughout the entire dissertation.

Finally, as a measure of relationship, the correlation between the intelligence quotient and the educational quotient was determined. The assumption underlying this device is that if pupils of normal intelligence are stimulated to work to capacity, then the correlation between intelligence and achievement will be high. To be sure, factors other than intelligence operate in these relationships, but numerous researches have proved beyond a doubt that intelligence is one of the important factors. In the second grade the differences existing between these relationships were very slight. The correlation coefficients of .28 ± .08 and .38 ± .07 for the undifferentiated and differentiated groups respectively show only a low positive relationship. These rather low coefficients indicate that the educational quotients cannot be predicted from the intelligence quotients in the second grade. It can be seen, therefore, that the difference of .1 ± .1 between the two coefficients as presented above is entirely insignificant because such, and much greater, differences could rightly be attributed to chance. For a graphical representation of these relationships the reader is referred to Figure 8 in the Appendix.

Summary of Second-Grade Findings

The more important findings for the second grade are those pertaining to the accomplishment ratio, the educational quotient, the composite score, and the correlation between the intelligence quotient and the educational quotient. The following items summarize the pertinent features of this portion of the study:

1. The undifferentiated group was superior to the differentiated group in educational accomplishment.

2. The undifferentiated group excelled in educational achievement as measured by the educational quotient.

3. The mean composite score for the undifferentiated group was 7.1 ± .9 higher than the mean composite score for the differentiated group.

Table I

Summary of Second-Grade Data

Item	Undifferentiated	Differentiated
Original Number of 2nd-Grade Pupils	89	64
Number of 2nd-Grade After Equating	61	61
Median I. Q. Before Equating	103.4	108.1
Median I. Q. After Equating	107.0	107.0
Modal Accomplishment Ratio	94.6	89.1
Mean Accomplishment Ratio	99.1 ± 1.4	90.9 ± 1.4
S. D. of Accomplishment Ratio	16.2 ± .1	15.6 ± .1
Skewness of Distribution	.3	.1
1st-Quartile Accomplishment Ratio[a]	89.1	78.8
Median Accomplishment Ratio	97.6	90.3
3rd-Quartile Accomplishment Ratio	107.1	99.2
Difference of Mean Accomplishment Ratios	8.2 ± 2.0	
Correlation Between I. Q.'s and E. Q.'s	.28± .08	.38± .07
Difference Between Coefficients of Correlation		.10± .1
1st-Quartile Educational Quotient[b]	95.8	83.5
Mean Educational Quotient	104.7 ± 1.1	95.2 ± 1.4
Difference of Mean Educational Quotients	9.5 ± 1.8	
3rd-Quartile Educational Quotient	114.8	105.8
Mean Composite Score	28.5 ± .7	21.4 ± .6
Difference of Mean Composite Scores	7.1 ± .9	
S. D. of Composite Score	8.0	7.5
Quartile Deviation of A. R.[c]	9.0	10.2
Quartile Deviation of E. Q.[d]	9.5	11.1

[a]Those having high intelligence quotients invariably had low accomplishment ratios.
[b]Educational quotients and intelligence quotients vary directly; hence the first quartile represents those of low intelligence.
[c]Accomplishment Ratio.
[d]Educational Quotient.

4. The undifferentiated group surpassed the differentiated group in educational achievement in the first-quartile, the median, and the third-quartile.

5. The low correlation between the intelligence quotients and the educational quotients reveals the fact that for second-grade children one cannot predict educational quotients from intelligence quotients.

6. The quartile deviation for the accomplishment ratios was 9.0 and 10.2 for the undifferentiated and differentiated respectively.[40]

7. The quartile deviation for the educational quotients was 9.5 and 11.1 for the undifferentiated and differentiated respectively. In both cases the greater variability existed in the differentiated group.

Third-Grade Findings

In the third-grade investigation the matching of the two groups started with ninety-four pupils in the undifferentiated group and eighty-three in the differentiated section. The final number in each group was sixty-six. This grade, like the second grade, had a normal intelligence quotient. Before and after matching the groups the median intelligence quotient for this grade was 109. The distributions conformed fairly well to the probability curve in every respect, showing a positive skewness with the distribution of the accomplishment ratio of .3 in each group.

The accomplishment ratios of the third-grade groups seemed to indicate the fact that no significant differences existed between the averages of the two groups. Only for the first quartile was the difference really significant. This rather interesting observation will take on meaning as the characteristics of this measure are more clearly understood. Before discussing the accomplishment ratio any further, several qualifying remarks may be necessary. Throughout this study it will be noted that those students having the highest intelligence quotients invariably had the lowest accomplishment ratios, while those of low intelligence had high accomplishment ratios. In the fifth grade the correlation between the intelligence quotient and the accomplishment ratio was — .40, thus showing that a negative relationship existed between intelligence and the accomplishment ratio. It must be understood, therefore, that the first-quartile accomplishment ratio was that of the highly intelligent pupils, while the third-quartile ratio represented those of low intelligence. The results of accomplishment in the third grade, then, definitely favored the highly intelligent in the undifferentiated group, but in both the median and the third-quartile divisions the difference on the side of the undifferentiated was negligible.

The educational quotient revealed nothing additional in favor of the undifferentiated group, but it very strikingly confirmed the assertion that the accomplishment ratio varies inversely while the educational quotient varies directly with the intelligence quotient. In this case, therefore, the greatest difference in favor of the undifferentiated group was found in the third-quartile, because for the educational quotient this represented the pupils of high intelligence just as the first quartile represented them in the accomplishment ratio.

When considering the mean composite score as measuring the difference of achievement, one sees quite clearly that the undifferentiated group was superior. It must be understood, however, that this fact alone becomes outstanding not because of the group as a whole but because of the excellent achievement of the upper-intelligence group. It is correct to say, then, with the above qualifications, that the undifferentiated group excelled in educational achievement.

The correlation between the intelligence quotient and the educational quotient was higher for the undifferentiated than for the differentiated group, but the difference was only $.22 \pm .09$ and could not be considered significant. It is interesting to note that while the coefficients were .28 for the undifferentiated group and .38 for the differentiated in the second grade, in the third grade the coefficients were .56 for the undifferentiated and only .34 for the differentiated. It appears, as a consequence of this fact alone, that all the improvement was made by the undifferentiated group. The correlations in the grades yet to be analyzed were higher for the undifferentiated in every grade except one, in which the coefficients were exactly equal. To be sure, this proves nothing for the problem of achievement directly, but it does point

[40] Toward the conclusion of this investigation it was deemed advisable to add to the description by including the quartile deviation as a measure of variability. The reader will note that the quartile deviations are mentioned only in the tables and in the summaries.

toward the conclusion that intelligence and achievement are more highly related to each other when pupils are not grouped by ability.

Summary of Third-Grade Findings

The conclusions for the third grade vary only slightly from those of the second grade, that is, they vary only in degree.

1. The undifferentiated group was superior to the differentiated group in the accomplishment ratio.

2. Using the educational quotient as a measure, the undifferentiated group excelled most noticeably in the third-quartile which represented the highly intelligent.

3. The mean composite score definitely favored the undifferentiated group, the difference being 5.6 ± 1.1 in its favor.

4. The correlation between the intelligence quotient and the educational quotient was higher for the undifferentiated group; the difference, however, between the two coefficients was negligible.

5. The higher coefficient of correlation for the third-grade undifferentiated group makes it possible to predict more accurately educational achievement from the intelligence quotients for the third grade than for the second grade; however, the coefficient for the differentiated group was too low to have much predictive value.

6. The quartile deviation for the undifferentiated accomplishment ratio was 7.4, for the differentiated it was 9.2. The quartile deviations for the educational quotients were 8.9 and 7.7 for the undifferentiated and the differentiated respectively.

Fourth-Grade Findings

The fourth-grade pupils of the two groups to be studied were quite difficult to match. From the original 132 pupils in the undifferentiated group and eighty in the differentiated group, the number dwindled to sixty-three for the equated groups. The reason for this difficulty lay with the original me-

Table II

Summary of Third-Grade Data

Item	Undifferentiated	Differentiated
Original Number of 3rd-Grade Pupils	94	83
Number of 3rd-Grade After Equating	66	66
Median I. Q. Before Equating	109.0	109.0
Median I. Q. After Equating	109.0	109.0
Modal Accomplishment Ratio	97.0	93.5
Mean Accomplishment Ratio	100.3 ± 1.	97.1 ± 1.
S. D. of Accomplishment Ratio	12.1 ± .7	11.7 ± .7
Skewness of Distribution	.3	.3
1st-Quartile Accomplishment Ratio[a]	92.9	87.0
Median Accomplishment Ratio	99.2	95.9
3rd-Quartile Accomplishment Ratio	107.6	105.3
Difference of Mean Accomplishment Ratios	3.2 ± 1.4	
Correlation Between I. Q.'s and E. Q.'s	.56± .06	.34± .07
Difference Between Coefficients of Correlation	.22± .09	
1st-Quartile Educational Quotient[b]	99.6	94.0
Mean Educational Quotient	107.6 ± 1.1	102.0 ± 1.0
Difference of Mean Educational Quotients	5.6 ± 1.5	
3rd-Quartile Educational Quotient	117.3	109.4
Mean Composite Score	44.8 ± .8	39.2 ± .8
Difference of Mean Composite Scores	5.6 ± 1.1	
S. D. of Composite Score	9.7	9.5
Quartile Deviation of A. R.[c]	7.4	9.2
Quartile Deviation of E. Q.[d]	8.9	7.7

[a]Those having high intelligence quotients invariably had low accomplishment ratios.
[b]Educational quotients and intelligence quotients vary directly; hence the first quartile represents those of low intelligence.
[c]Accomplishment Ratio.
[d]Educational Quotient.

dian intelligence quotient. Inasmuch as the original intelligence quotient for the undifferentiated group was 95.3 as against that of 106.5 for the differentiated group, considerable culling became necessary in the undifferentiated group to balance the difference of 11.2 points in their intelligence. It is interesting to note that in every grade except the third the original median intelligence quotient was higher for the differentiated group than for the undifferentiated, but in no case was the difference as great as that which existed in the fourth grade. No reason can be given here for this variation. It would seem according to available evidence that the schools practicing differentiation are those having the pupils of higher intelligence. After the process of matching in the fourth-grade groups had been effected, the median intelligence quotient for both was 102.5, which quotient is slightly lower than the median intelligence quotients considered before.

The frequency distributions of accomplishment ratios did not show the same type of symmetry in the fourth grade as that noticed in the first and the second grades. Both distributions were skewed negatively, the skewness being — .5 for the undifferentiated group and — .1 for the differentiated. Since a coefficient of skewness equal to zero characterizes a perfectly symmetrical distribution, it is evident, then, that the greatest amount of symmetry existed in the differentiated group. Another difference occurring in the distribution was that of variability. The standard deviation of the undifferentiated group was 7.5 ± .5, but that of the differentiated group was 11.2 ± .7. With these factors before us, it is possible to examine the educational achievement of the fourth grade.

The accomplishment ratios of the two groups revealed very little difference in favor of any group. However, some noticeable difference existed in the third quartile which, it must be remembered, represented a low intelligence level. The undifferentiated group had a third-quartile accomplishment ratio of 107.8 as compared with that of 112.6 for the differentiated group. The difference between the means of the two groups was entirely negligible. According to the accomplishment ratios, there was a slight advantage for the differentiated of low intelligence in the third quartile.

Another new item of interest resulted from the analysis of the fourth grade in connection with the educational quotient. In the two grades analyzed before, there was definite agreement between the accomplishment ratio and the educational quotient, but in the fourth grade there was almost perfect disagreement in the third quartile. As mentioned before, the third-quartile educational quotient represented the high intelligence level, just as the first quartile accomplishment ratio represented the low level. This being true, it would seem that these values should vary similarly. If it is remembered, however, that the educational quotient is the value obtained by dividing the educational age by the chronological age, it is reasonable to assume that the brilliant can gain in educational age without making comparable improvement in the accomplishment ratio.

The mean composite scores for both groups were almost exactly equal, the difference being only 1.3 ± 1.2. The standard deviations of the composite score distributions did not reveal the differences in dispersion noticed in the accomplishment ratio. This further substantiates the fact that the differences mentioned in the educational quotients were due to factors other than scores.

The coefficients of correlation for the fourth grade were .75 ± .04 and .61 ± .05 for the undifferentiated and differentiated groups respectively. It would, therefore, be possible, with a fair degree of accuracy, to predict educational achievement by means of the intelligence quotients. Such predictive possibilities were not evident in either the second or the third grades, but the difference between the coefficients of the fourth grade groups was too small to conclude that the relationship between them makes possible to any very significant degree greater prediction for the one than for the other. The difference of .14 ± .06 between the coefficients of the two groups was not a statistically significant difference.

Summary of Fourth-Grade Findings

The conclusions in summary form for the fourth grade are as follows:

1. A considerable difference, existing before the matching of the two groups, necessitated the throwing out of many undifferentiated cases.

2. The median intelligence quotient after matching was only 102.5, which is lower than both the second and the third grade intelligence quotients.

3. Both groups were approximately equal in accomplishment, the greatest difference being in the third quartile. The difference existing between the third quartile accomplishment ratios was 4.8, in favor of the differentiated group.

4. The differentiated group was slightly superior to the undifferentiated with respect to the educational quotient, the greatest difference being in the third quartile.

5. Both groups were approximately equal in the mean composite scores.

6. The correlation between the intelligence quotient and the educational quotient shows that for the fourth grade it is possible to predict rather accurately the educational quotients from the intelligence quotients.

7. The quartile deviations for the accomplishment ratios were 5.6 and 8.3 for the undifferentiated and the differentiated respectively. For the educational quotients the deviations were 8.0 and 9.8 respectively.

Fifth-Grade Findings

The fifth-grade groups in their original tabulation exhibited some of the same difficulties as those of the fourth grade. From 154 undifferentiated pupils and ninety-four differentiated ones the matching process brought both groups to eighty-three in number. The median intelligence quotients prior to the equating were 94.7 for the undifferentiated group and 103.0 for the differentiated. The equating process brought the median intelligence quotient to 100.3 for both groups. Although 100.3 is considered a normal intelligence quotient, it is somewhat lower than any other thus far considered. The undifferentiated group presented a distribution in the accomplishment ratios which conformed almost exactly to the probability curve, the skewness being only .07. The differentiated group had a positive skewness of .3. The dispersion of the differentiated group was almost twice that of the undifferentiated, with a standard deviation of 8.8 ± .5 for the un-

Table III

Summary of Fourth-Grade Data

Item	Undifferentiated	Differentiated
Original Number of Fourth-Grade Pupils	132	80
Number of 4th-Grade After Equating	63	63
Median I. Q. Before Equating	95.3	106.5
Median I. Q. After Equating	102.5	102.5
Modal Accomplishment Ratio	105.9	105.8
Mean Accomplishment Ratio	102.0 ± .7	104.3 ± .9
S. D. of Accomplishment Ratio	7.6 ± .5	11.2 ± .7
Skewness of Distribution	—.5	—.1
1st-Quartile Accomplishment Ratio[a]	96.7	96.7
Median Accomplishment Ratio	103.3	104.8
3rd-Quartile Accomplishment Ratio	107.8	112.6
Difference of Mean Accomplishment Ratios		2.3 ± 1.1
Correlation Between I. Q.'s and E. Q.'s	.75 ± .04	.61 ± .05
Difference Between Coefficients of Correlation	.14 ± .06	
1st-Quartile Educational Quotient[b]	95.2	96.3
Mean Educational Quotient	102.9 ± .9	105.8 ± 1.1
Difference of Mean Educational Quotients		2.9 ± 1.4
3rd-Quartile Educational Quotient	111.8	115.9
Mean Composite Score	55.2 ± .9	56.5 ± .8
Difference of Mean Composite Scores		1.3 ± 1.2
S. D. of Composite Score	10.2	10.0
Quartile Deviation of A. R.[c]	5.6	8.3
Quartile Deviation of E. Q.[d]	8.0	9.8

[a]Those having high intelligence quotients invariably had low accomplishment ratios.
[b]Educational quotients and intelligence quotients vary directly; hence the first quartile represents those of low intelligence.
[c]Accomplishment Ratio.
[d]Educational Quotient.

differentiated and 15.7 ± .8 for the differentiated in accomplishment ratios.

The accomplishment ratios of the two groups showed no significant difference between the groups when their averages were considered. When, on the other hand, each quartile was studied separately one observed that the first-quartile ratios were exactly equal, the median ratio was only slightly higher for the differentiated group, and the third-quartile ratio favored the differentiated group. The variation throughout was very slight. It was easy to conclude, therefore, that both groups were approximately equal in educational accomplishment.

The writer wishes here to justify his position in using, not one measure to describe the educational status, but the four measures discussed earlier. Had this four-fold technique[41] not been used, the status of the fourth-grade could never have been accurately described, nor could the fifth-grade observations have been adequately determined. If the reader were to reach conclusions from the evidence presented, he would, through the application of the first technique, naturally say that both grades were equal in educational achievement. Such an inference would be incorrect. Achievement, dependent as it is upon many factors, can never be accurately described through only one particular measure.

The educational quotient, however, adds very little color to the educational-achievement picture. Upon examining the various levels, the first-quartile, the median, and the third-quartile, the observer will note that the only existing difference worthy of consideration herewith is that which existed in the third-quartile, wherein the differentiated group excelled. Such a difference is quite interesting, especially when considered in the light of the mean composite scores. The mean composite scores thrust a somewhat contradictory item into our analysis, for it is precisely at this point that the balance shifted in favor of the differentiated group. The difference between the mean composite scores of 5.9 ± 1.2 was too great to be attributable to chance factors. The two groups achieved approximately the same educational-age status, as determined by the educational quotient. When the actual age, the educational age, and the educational accomplishment are considered variables, there exists the possibility of one group having a higher composite score than the other group. From this standpoint, then, the differentiated group was superior only in so far as the composite score alone was concerned.

It is important to note here especially the correlation between the intelligence quotient and the educational quotient, for such a correlation reaches its high point in the fifth-grade undifferentiated group. As noticed before, the coefficients of correlation gradually rose in value, as did also the difference between the two groups. At first the values were low and the differences entirely insignificant; but later values increased and significant differences arose. The coefficients of correlation were .79 ± .03 and .60 ± .05 for the undifferentiated and differentiated groups, respectively, the difference being .19 ± .05. For the fifth grade, where the difference was more than three times its probable error, the difference is of doubtful significance.

Summary of Fifth-Grade Findings

The important findings for the fifth grade are as follows:

1. Both groups were essentially equal in accomplishment and educational quotient.
2. The differentiated group was decidedly superior in the composite score.
3. The correlation between the intelligence quotient and the educational quotient reached its high point in the fifth-grade undifferentiated groups.
4. The difference between the coefficients of correlation was .19 ± .05 in favor of the undifferentiated group.
5. The undifferentiated quartile deviations are 5.8 for the accomplishment ratio and 9.2 for the educational quotient, while the differentiated deviations are 6.3 and 11.2 for accomplishment and educational quotient respectively.

Sixth-Grade Findings

In the sixth-grade groups, difficulties again presented themselves as they did in the fourth and fifth grades. Originally there were 154 pupils in the undifferentiated group and ninety-four in the differentiated. That the pupils were quite easily matched may be seen

[41] The accomplishment ratio, the educational quotient, the mean composite score, and the correlation between the intelligence quotient and the educational quotient constitute the factors in the four-fold technique.

TABLE IV

SUMMARY OF FIFTH-GRADE DATA

Item	Undifferentiated	Differentiated
Original Number of 5th-Grade Pupils	154	94
Number of 5th Grade After Equating	83	83
Median I. Q. Before Equating	94.7	103.0
Median I. Q. After Equating	100.3	100.3
Modal Accomplishment Ratio	102.1	100.3
Mean Accomplishment Ratio	102.7 ± .6	105.7 ± 1.2
S. D. of Accomplishment Ratio	8.8 ± .5	15.7 ± .8
Skewness of Distribution	.07	.3
1st-Quartile Accomplishment Ratio[a]	97	97
Median Accomplishment Ratio	102.5	103.9
3rd-Quartile Accomplishment Ratio	108.5	109.5
Difference of Mean Accomplishment Ratios		3.0 ± 1.3
Correlation Between I. Q.'s and E. Q.'s	.79± .03	.60± .05
Difference Between Coefficients of Correlation	.19± .05	
1st-Quartile Educational Quotient[b]	93.1	91.4
Mean Educational Quotient	101.9 ± .9	103.9 ± 1.0
Difference of Mean Educational Quotients		2.0 ± 1.3
3rd-Quartile Educational Quotients	111.5	115.7
Mean Composite Score	65.5 ± .8	71.4 ± .9
Difference of Mean Composite Scores		5.9 ± 1.2
S. D. of Composite Score	11.0	13.0
Quartile Deviation of A. R.[c]	5.8	6.3
Quartile Deviation of E. Q.[d]	9.2	11.2

[a]Those having high intelligence quotients invariably had low accomplishment ratios.
[b]Educational quotients and intelligence quotients vary directly; hence the first quartile represents those of low intelligence.
[c]Accomplishment Ratio.
[d]Educational Quotient.

in the number that finally remained. The number of "throwouts" was only ten in the differentiated, but seventy in the undifferentiated group, thus leaving a total of eighty-four for the matched groups. The two groups differed widely in their original intelligence status. The median intelligence quotient was ninety-five for the undifferentiated group and 105.9 for the differentiated. The process of matching changed it for both groups to 104.1, which was slightly better than the medians for the fourth and fifth grades. The distributions for the accomplishment ratios were fairly symmetrical for both groups. As in the two grades just analyzed the dispersion in the differentiated group was approximately twice that of the undifferentiated. The dispersion, as measured by the standard deviation, was 7.8 ± .4 and 14.2 ± .7 for the undifferentiated and the differentiated groups respectively. These facts made it possible to analyze the two groups according to the method employed in the earlier grades.

The two groups showed some difference at every level in their educational accomplishment. These differences became consistently greater from the first quartile to the third. In the first quartile the difference was only 2.1 in favor of the differentiated group, while in the median and the third quartile the differences in favor of the differentiated were 3.5 and 6.8 respectively. The difference between the mean accomplishment ratios was 4.1 ± 1.2, a difference which is barely significant except as it favored the differentiated group. The assertion, therefore, holds that the differentiated group excels slightly in educational accomplishment, the greatest difference being in the third quartile which represents the accomplishment of pupils having low intelligence quotients.

The educational quotients of the sixth grade showed results almost identical with those obtained in the fifth grade. In the first quartile the difference was quite negligible; in the means, a difference of 4.8 ± 1.4 in favor

of the differentiated group verified an appreciable difference, though still slightly less than four times its probable error; and in the third quartile a difference of 8.7 favored the differentiated group. Here again the bright pupils in the differentiated groups excelled more significantly in their educational quotients than in their educational accomplishment. It appears, therefore, when their educational achievement is compared with established norms, that bright pupils usually do fairly well in school. When, however, we compare what they do with what they can do, measuring what they can do by their intelligence quotient, we discover a regretable deficiency in their achievement. According to the results obtained, there was practically no evidence that grouping prevented such deficiency, although there could be no inference that grouping could not prevent it. It should be noted here, prior to further analysis, that before pupils can excel in every measure they must work to full capacity.

The application of the third measure, referred to as the composite score, showed a difference of 4.1 ± 1.1 in favor of the differentiated group. Such a difference, when considered in the light of the results obtained by means of the other two measures, cannot be ignored but must be considered as definitely favoring the differentiated group. This fact substantiates the assertion previously made that high scores do not necessarily produce high accomplishment ratios, for in themselves they do not always show the relation between achievement and ability; nor do they guarantee high educational quotients, in as much as they do not take into account the age of the pupil.

The correlation between the intelligence quotient and the educational quotient presented a very contradictory item for analysis.[42] It was here that the difference between the coefficients of correlation of $.26 \pm .05$ definitely favored the undifferentiated group. The undifferentiated pupils of the sixth grade showed a greater degree of relationship between intelligence and achievement, according to the measure of correlation used, than did the differentiated, the correlation coefficients being $.79 \pm .03$ and $.54 \pm .05$ for the undifferentiated and the differentiated, respectively.

Summary of Sixth-Grade Findings

The following findings reveal the educational status of the two groups for the sixth grade.

1. The differentiated group was superior in educational accomplishment. The greatest difference occurred in the third quartile, representing the pupils of low intelligence.

2. The differentiated group excelled in the educational quotient, the greatest difference existing in the third quartile representing the pupils of low intelligence.

3. The differentiated group had a higher mean composite score. The difference between the two groups was 4.1 ± 1.1, in favor of the differentiated group.

4. A high composite score does not necessarily produce a correspondingly high accomplishment ratio, or educational quotient.

5. The undifferentiated group was definitely superior to the differentiated group, according to the correlation between the intelligence quotient and the educational quotient. The difference between the two quotients was $.25 \pm .05$ in favor of the undifferentiated group.

6. The quartile deviations again were smaller for the undifferentiated group; they were 4.4 for accomplishment and 8.6 for the educational quotients, while the deviations for the differentiated were 8.6 and 12.7 for accomplishment and educational quotients.

Seventh-Grade Findings

The seventh-grade groups studied in this investigation embraced more pupils than any other groups in the area under consideration. From the original number of 180 in the undifferentiated group and 298 in the differentiated, the final groups containing 171 pupils were selected and matched. The median intelligence quotients of the unmatched groups were 97.5 and 105.8 for the undifferentiated and differentiated respectively. These medians, rather low before they were matched, were predictive of the low matched medians which in their final analysis were 98.6 for both groups. This rather low median intelligence quotient necessitated the elimination of many of the highest-ability pupils in both groups, thus affecting some of the other measures to be considered later.

[42] Superiority of any one group does not necessarily guarantee prediction possibilities for educational quotients from intelligence quotients.

TABLE V

SUMMARY OF SIXTH-GRADE DATA

Item	Undifferentiated	Differentiated
Original Number of 6th-Grade Pupils	154	94
Number of 6th-Grade After Equating	84	84
Median I. Q. Before Equating	95.0	105.9
Median I. Q. After Equating	104.1	104.1
Modal Accomplishment Ratio	102.7	104.8
Mean Accomplishment Ratio	102.1 ± .6	106.3 ± 1
S. D. of Accomplishment Ratio	7.8 ± .4	14.2 ± .7
Skewness of Distribution	—.07	+.1
1st-Quartile Accomplishment Ratio[a]	97.8	99.9
Median Accomplishment Ratio	102.3	105.8
3rd-Quartile Accomplishment Ratio	106.5	113.3
Difference of Mean Accomplishment Ratios		4.2 ± 1.2
Correlation Between I. Q.'s and E. Q.'s	.79± .03	.54± .05
Difference Between Coefficients of Correlation	.25± .05	
1st-Quartile Educational Quotient[b]	96.5	97.0
Mean Educational Quotient	104.6 ± .9	109.4 ± 1.1
Difference of Mean Educational Quotients		4.8 ± 1.4
3rd-Quartile Educational Quotient	113.6	122.3
Mean Composite Score	78.0 ± .7	82.1 ± .9
Difference of Mean Composite Scores		4.1 ± 1.1
S. D. of Composite Score	10.0	11.7
Quartile Deviation of A. R.[c]	4.4	6.7
Quartile Deviation of E. Q.[d]	8.6	12.7

[a]Those having high intelligence quotients invariably had low accomplishment ratios.
[b]Educational quotients and intelligence quotients vary directly; hence the first quartile represents those of low intelligence.
[c]Accomplishment Ratio.
[d]Educational Quotient.

The accomplishment ratios show very little difference between the two groups. Their respective distributions showed a skewness of zero for the undifferentiated and —.3 for the differentiated group. The three levels of educational accomplishment were approximately equal in every respect. The greatest difference, occurring in the first quartile, amounted to only two points in favor of the undifferentiated group. The slight and insignificant difference between the means of .77 ± 1.1 on the side of the undifferentiated may be waived entirely.

The educational quotients revealed no additional evidence in favor of any group. They merely verified the conclusion that no real difference could be found between the two groups. At only one level, the first quartile, did the difference amount to as much as 4.7 points in favor of the undifferentiated group. The difference between the mean educational quotients of 2.1 ± 1.1 in favor of the undifferentiated must be considered negligible.

The mean composite scores again substantiated the fact that both groups were essentially equal. These means were equal to 82.6 ± .5 and 80.9 ± .6 for the undifferentiated and the differentiated respectively, giving an insignificant difference of 1.7 ± .8 on the side of the undifferentiated.

The correlation between the intelligence quotients and the educational quotients was essentially equal for both groups. It is well to note in passing that here, for the first time in the experiment, the two correlation coefficients were almost identical. The coefficients were .76 ± .02 for the undifferentiated and .74 ± .02 for the differentiated. Both of these, indicating as they do a high degree of relationship, assist in establishing the fact that, for the seventh-grade pupils, one can safely predict the educational quotients from the intelligence quotients. The investigator has no authority to claim that this sudden rise in relationship on the part of the differentiated group was due in any

way to a particular school organization, but the reader should notice that the rise occurred simultaneously with the change to the junior high school set up. It so happened that for this study the differentiated pupils for grades seven, eight, and nine were all found in systems having a junior high school. Further study of this problem would be justified.

Summary of Seventh-Grade Findings

The summary of the seventh-grade data involves the following conclusions:

1. Both groups were essentially equal in educational accomplishment.

2. The two groups were about equal in educational quotients, the only real difference occurring in the first quartile and in favor of the undifferentiated group.

3. The mean composite scores revealed a difference of 2.1 ± 1.1 in favor of the undifferentiated group. So small a difference cannot be considered significant.

4. The coefficients of correlation between intelligence quotients and educational quotients were just about equal. The sudden rise in relationship on the part of the differentiated group occurred simultaneously with the first year of the junior high school organization.

5. The quartile deviations in accomplishment were 6.3 and 7.0 for the undifferentiated and differentiated respectively, while in educational quotients the deviations were 10.0 and 11.3 for the two groups respectively.

Eighth-Grade Findings

Prior to the process of matching there were 146 pupils in the undifferentiated group and 278 in the differentiated. The number was reduced to 121 for each group by matching. The original median intelligence quotients were 99.3 and 109.8 for the undifferentiated and differentiated respectively.

The median intelligence quotient after equating was 102.4. It is interesting to note

Table VI

Summary of Seventh-Grade Data

Item	Undifferentiated	Differentiated
Original Number of 7th-Grade Pupils	180	298
Number of 7th-Grade After Equating	171	171
Median I. Q. Before Equating	97.5	105.8
Median I. Q. After Equating	98.6	98.6
Modal Accomplishment Ratio	101.3	101.4
Mean Accomplishment Ratio	101.3 ± .5	99.6 ± .3
S. D. of Accomplishment Ratio	10.8 ± .4	5.8 ± .2
Skewness of Distribution	0	—.3
1st-Quartile Accomplishment Ratio[a]	94.9	92.9
Median Accomplishment Ratio	101.3	100.2
3rd-Quartile Accomplishment Ratio	107.4	106.9
Difference of Mean Accomplishment Ratios	.77 ± 1.1	
Correlation Between I. Q.'s and E. Q.'s	.76 ± .02	.74 ± .02
Difference Between Coefficients of Correlation	.02 ± .03	
1st-Quartile Educational Quotient[b]	89.9	85.2
Mean Educational Quotient	98.6 ± .78	96.4 ± .8
Difference of Mean Educational Quotients	2.1 ± 1.1	
3rd-Quartile Educational Quotient	109.8	107.8
Mean Composite Score	82.6 ± .5	80.9 ± .6
Difference of Mean Composite Scores	1.7 ± .8	
S. D. of Composite Score	9.0	12.5
Quartile Deviation of A. R.[c]	6.3	7.0
Quartile Deviation of E. Q.[d]	10.0	11.3

[a]Those having high intelligence quotients invariably had low accomplishment ratios.
[b]Educational quotients and intelligence quotients vary directly; hence the first quartile represents those of low intelligence.
[c]Accomplishment Ratio.
[d]Educational Quotient.

that equating improved the status of the undifferentiated group. Such improvement resulted quite frequently from the matching process, because in practically every case the original differentiated group was superior in intelligence to the undifferentiated. The superiority of the differentiated group over the second group always necessitated the throwing out of the scores representing the low intelligence level in the undifferentiated group. Another factor to notice, rather than to explain, is the continual shifting in median intelligence quotients. The reader will remember that grade seven was exceedingly low in intelligence, despite the fact that seventh-grade pupils are maturer in years and experience. This being true, several changes were to be expected in the accomplishment ratios and the educational quotients.

The accomplishment ratios varied only a little, but they favored the undifferentiated group more than did the ratios in the seventh grade. At every level the undifferentiated group was slightly superior. However, the difference between the mean ratios was only $2.6 \pm .8$ and could not be considered very striking. It must be remembered that the undifferentiated group did not excel greatly at any one level but rather slightly at every level. Consequently, the group as a whole proved itself definitely superior in educational accomplishment.

The educational quotients were affected considerably by the increase in the median intelligence quotient. While in the accomplishment ratio only a slight decrease was evident, in the educational quotients a definite increase was noticeable. Here, again, it became obvious that a rise in intelligence produces a decrease in the accomplishment ratio but, on the contrary, a corresponding increase in the educational quotient. Another feature about the analysis of the educational quotient manifested itself in the fact that, like the accomplishment ratio, it was higher at every level of intelligence. Since the difference between the means, however, was only 2.5 ± 1.3 there was nothing to argue superiority for the undifferentiated intelligence quotients.

The greatest evidence in favor of the undifferentiated group became noticeable at this point in the mean composite scores. The mean composite scores for the eighth grade were higher than for any other grade previously analyzed, and thus coincided with the normal achievement of that grade. The difference between the mean composite scores very definitely favored the undifferentiated group. The difference between $99.3 \pm .6$ for the undifferentiated and $90.5 \pm .7$ for the differentiated was $8.8 \pm .9$ in favor of the undifferentiated pupils. The correlation between the intelligence quotients and the educational quotients was exactly the same for both groups. The coefficients for both groups were equal to $.79 \pm .02$. Such relationship can rightly be considered high.

Summary of Eighth-Grade Findings

The findings in this grade are similar to those of the seventh grade but more especially favorable to the undifferentiated group.

1. The undifferentiated group was slightly superior in educational accomplishment.
2. The difference between the mean educational quotients favored the undifferentiated group with a small significance.
3. The undifferentiated group excelled in the mean composite scores, the difference being $8.8 \pm .9$ in its favor.
4. The high coefficients of correlation between the intelligence quotients and the educational quotients substantiated a contention of this investigation that it is possible to predict educational quotients with a fair degree of accuracy.
5. The quartile deviations were 6.5 and 7.0 for the undifferentiated and differentiated in accomplishment. For the educational quotients the deviations were 10.7 and 11.2 for the undifferentiated and differentiated.

Ninth-Grade Findings

The ninth grade was the last to be analyzed and in general the most interesting one. There were 214 pupils in the undifferentiated group and 171 in the differentiated prior to the matching of the two groups. The remaining number in each group after the matching was 138. Before matching the pupils the median intelligence quotients were 104.4 and 110.9 for the undifferentiated and the differentiated respectively. It will be noticed that the quotient was unusually high for the differentiated pupils. The median for the matched groups was 107.3. This median quotient was higher than that of all the other grades, excepting the second and the third grades.

The accomplishment ratios presented frequency distributions that were essentially the

TABLE VII

SUMMARY OF EIGHTH-GRADE DATA

Item	Undifferentiated	Differentiated
Original Number of 8th-Grade Pupils	146	278
Number of 8th-Grade After Equating	121	121
Median I. Q. Before Equating	99.3	109.8
Median I. Q. After Equating	102.4	102.4
Modal Accomplishment Ratio	99.8	99.2
Mean Accomplishment Ratio	101.3 ± .5	98.6 ± .6
S. D. of Accomplishment Ratio	8.4 ± .4	10.0 ± .4
Skewness of Distribution	.2	—.1
1st-Quartile Accomplishment Ratio[a]	94.5	92.0
Median Accomplishment Ratio	100.8	98.8
3rd-Quartile Accomplishment Ratio	107.5	106.0
Difference of Mean Accomplishment Ratios	2.6 ± .8	
Correlation Between I. Q.'s and E. Q.'s	.79 ± .02	.79 ± .02
Difference Between Coefficient of Correlation	0	0
1st-Quartile Educational Quotient[b]	91.8	88.9
Mean Educational Quotient	102.9 ± .8	100.4 ± 1.0
Difference of Mean Educational Quotients	2.5 ± 1.3	
3rd-Quartile Educational Quotient	113.1	111.3
Mean Composite Score	99.3 ± .6	90.5 ± .7
Difference of Mean Composite Scores	8.8 ± .9	
S. D. of Composite Score	9.4	11.1
Quartile Deviation of A. R.[c]	6.5	7.0
Quartile Deviation of E. Q.[d]	10.7	11.2

[a]Those having high intelligence quotients invariably had low accomplishment ratios.
[b]Educational quotients and intelligence quotients vary directly; hence the first quartile represents those of low intelligence.
[c]Accomplishment Ratio.
[d]Educational Quotient.

same in skewness and dispersion. The skewness for the undifferentiated was — .3 and for the differentiated — .2. The standard deviations were 7.8 ± .3 and 8.0 ± .3 for the undifferentiated and the differentiated respectively. At every level the differentiated group excelled in educational accomplishment; the difference between the means was 3.0 ± .6 in favor of the differentiated group. This difference was five times its probable error and very definitely favorable to the differentiated group.

The educational quotients presented no striking differences between the two groups. The difference between the mean quotients was 1.3 ± 1.0, hence entirely insignificant. This insignificance needs to be explained, because the casual observer would consider it a virtual impossibility. Remembering, however, that a high educational age and a high accomplishment ratio are two quite independent factors, one finds little difficulty in understanding the possibility of these apparently incongruent ratings of the two ninth-grade groups.

The undifferentiated and the differentiated groups of the ninth grade exhibited striking differences in their composite scores. The mean composite score of the undifferentiated group was 98.7 ± .6. This was slightly lower than the mean score of the undifferentiated eighth-grade group. The mean composite score of the differentiated group, however, was raised from 90.5 ± .7 of the eighth grade to 102.3 ± .4 in the ninth. This latter score is the highest mean score attained in this investigation. One can see, then, that the difference of 3.6 ± .7 favoring the differentiated group in the ninth grade is statistically significant.

The correlation between the intelligence quotients once more establishes the fact that for both groups it is possible to predict the educational quotients from the intelligence quotients to about the same degree of accuracy. The coefficients of correlation were .76 ± .02 and .71 ± .03 for the undifferentiated and the differentiated groups respectively.

TABLE VIII

SUMMARY OF NINTH-GRADE DATA

Item	Undifferentiated	Differentiated
Original Number of 9th-Grade Pupils	214	171
Number of 9th-Grade After Equating	138	138
Median I. Q. Before Equating	104.4	110.9
Median I. Q. After Equating	107.3	107.3
Modal Accomplishment Ratio	99.2	101.3
Mean Accomplishment Ratio	97.1 ± .4	100.1 ± .5
S. D. of Accomplishment Ratio	7.8 ± .3	8.0 ± .3
Skewness of Distribution	—.3	—.2
1st-Quartile Accomplishment[a]	92.3	94.1
Median Accomplishment Ratio	97.8	100.5
3rd-Quartile Accomplishment Ratio	102.5	104.9
Difference of Mean Accomplishment Ratios		3.0 ± .6
Correlation Between I. Q.'s and E. Q.'s	.76± .02	.71± .03
Difference Between Coefficients of Correlation	.05± .04	
1st-Quartile Educational Quotient[b]	94.3	102.5
Mean Educational Quotient	104.6 ± .8	105.9 ± .6
Difference of Mean Educational Quotients		1.3 ± 1.0
3rd-Quartile Educational Quotient	115.2	113.0
Mean Composite Score	98.7 ± .6	102.3 ± .4
Difference of Mean Composite Scores		3.6 ± .7
S. D. of Composite Score	10.2	7.1
Quartile Deviation of A. R.[c]	5.1	5.4
Quartile Deviation of E. Q.[d]	10.5	5.3

[a]Those having high intelligence quotients invariably had low accomplishment ratios.
[b]Educational quotients and intelligence quotients vary directly; hence the first quartile represents those of low intelligence.
[c]Accomplishment Ratio.
[d]Educational Quotient.

TABLE IX

SUMMARY OF GRADES TWO TO NINE INCLUSIVE

Item	Undifferentiated	Differentiated
Original Number in Study	1163	1162
Number After Equating	787	787
Mean I. Q. After Equating	103.1	103.1
Mean E. Q. After Equating	102.9	102.0
Difference of Mean Accomplishment Ratios	.2 ± .38	
Mean Accomplishment Ratio	100.6 ± .24	100.4 ± .30
S. D. of Accomplishment Ratio	10.2	12.3
Correlation Between I. Q.'s and E. Q.'s	.73± .01	.61± .015
Difference Between Coefficients of Correlation	.12± .018	
1st-Quartile Educational Quotient	93.5	91.5
Mean Educational Quotient	102.9 ± .3	102.0 ± .4
Difference of Mean Educational Quotients	.9 ± .5	
3rd-Quartile Educational Quotient	113.3	112.1
Mean Composite Score	76.2	75.2
Difference of Mean Composite Scores	1.0	
Quartile Deviation of Educational Quotients	9.9	10.3

Summary of Ninth-Grade Findings

The findings of the ninth grade decidedly favor the differentiated group; they are as follows:

1. The differentiated group was decidedly superior in educational accomplishment.

2. The two groups were approximately equal in the educational quotients, the greatest difference existing in the first quartile which represented the pupils of low intelligence.

3. The differentiated group excelled the undifferentiated in the composite scores. The difference between the means was $3.6 \pm .7$.

4. The high degree of relationship between the intelligence quotients and the educational quotients was essentially the same in both groups, the coefficients being $.76 \pm .02$ and $.71 \pm .03$ for the undifferentiated and the differentiated respectively.

5. The quartile deviations were 5.1 and 5.4 for the undifferentiated and differentiated in accomplishment. The quartile deviations of the educational quotients were 10.5 and 5.3 for the two groups respectively.

Findings of Grades Two to Nine Inclusive

The total number of pupils in the entire range of the grades under study was 1,163 in the undifferentiated groups and 1,162 in the differentiated groups. By the process of matching, this number was reduced to 787 pairs for both groups. The difference between 1,163 and 787 represented the number lost by matching the undifferentiated groups, and the difference between 1,162 and 787, the number lost by matching the differentiated groups. Both groups presented an average distribution of intelligence, the matched mean intelligence quotient being 103.1.

The educational accomplishment of the two groups, as measured by the accomplishment ratio, was almost identical for both, the difference being only $.2 \pm .38$ in favor of the undifferentiated group. This difference, entirely insignificant, carries no weight and is attributable to chance factors. When the individual grades are considered, however, this result can be described in still another manner. In two grades the undifferentiated pupils were superior in their mean accomplishment ratios, but in two grades also the differentiated were to much the same extent superior. Grades two and eight excelled in the undifferentiated means, and grades six and nine were superior in the differentiated means. The other four grades were also equally divided, grades three and seven presenting only a slight measurable advantage for the undifferentiated, and grades four and five favoring the differentiated to about the

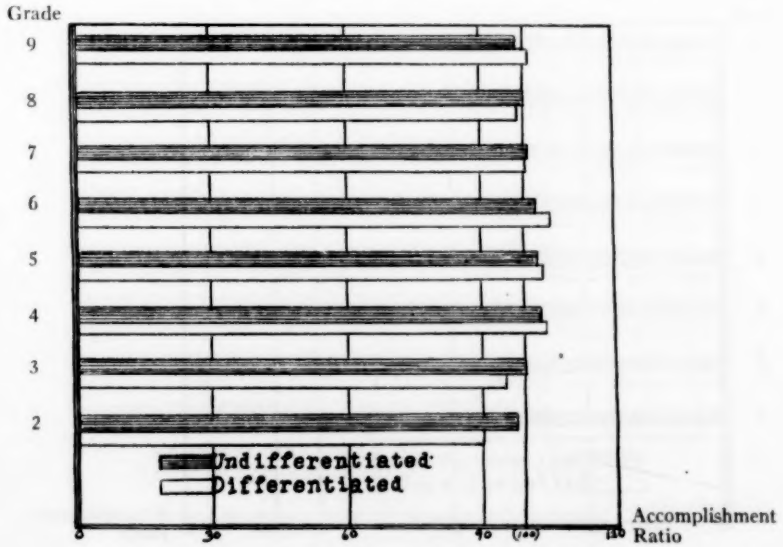


Figure 1.—Median Accomplishment Ratios.

same extent. Table X presents the statistical evidence in favor of the various groups and grades.

TABLE X

MEAN DIFFERENCE OF ACCOMPLISHMENT RATIOS

Grade	Undifferentiated	Differentiated
All	.2 ± .38	
9		3.0 ± .6
8	2.6 ± .8	
7	.77 ± 1.1	
6		4.2 ± 1.2
5		3.0 ± 1.3
4		2.3 ± 1.1
3	3.2 ± 1.4	
2	8.2 ± 2.0	

The central tendency, as measured by the median accomplishment ratios, presents the same evidence as that of the means. This evidence is presented graphically in Figure 1.

The first-quartile accomplishment ratios, representing generally those pupils of high intelligence, revealed similarities with those of the mean and the median. The slight measurable advantage in favor of the differentiated group in the fourth and the fifth grades, however, completely disappeared, leaving a difference in favor of the differentiated group in grades six and nine. The undifferentiated retained their superiority in grades two, three, seven, and eight for both the first quartile measure and the two central-tendency measures. Grades five and six were exactly equal in both groups. Table XI shows the values attained by each group for the various grades. The starred values indicate which group was superior and in what particular grade it was superior.

TABLE XI

FIRST-QUARTILE ACCOMPLISHMENT RATIOS

Grade	Undifferentiated	Differentiated
9	92.3	94.1*
8	94.5*	92.0
7	94.9*	92.9
6	97.8	99.9*
5	97.0	97.0
4	96.7	96.7
3	92.9*	87.0
2	89.1*	78.8

*Indicates superiority, which when compared will indicate what group attained the higher ratio at the point exceeding twenty-five percent of all the ratios.

Figure 2 presents the evidence of Table XI in a graphical form and facilitates accomplish-

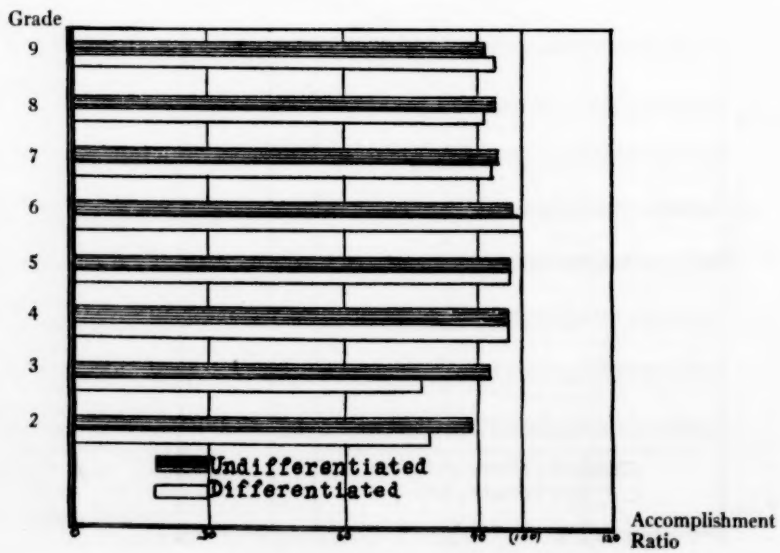


Figure 2.—First-Quartile Accomplishment Ratios.

ment evaluation by noting the one hundred line. An examination of the various graphs will reveal the relative distances existing for the various accomplishment levels from the one hundred line. This same evidence, however, can be deduced from the tabular presentation.

For the third-quartile level the undifferentiated maintained their superiority in grades two, three, seven, and eight, while the differentiated excelled in grades four, five, six, and nine. The outstanding difference in accomplishment appeared in the fourth grade; in this one case the differentiated were decidedly superior, although for the mean the difference was only slight, and for the first quartile none at all. Table XII presents the values for the third quartile and permits easy comparison with Tables X and XI. (See Figure 3 for the graphical presentation of the same.)

In the concluding analysis of the accomplishment ratios, it may be seen that viewed as a whole both groups were virtually equal; but when considered grade by grade, grades two, three, seven, and eight were superior in the undifferentiated groups at every level, and grades six and nine excelled throughout in the differentiated groups. Grades four and five showed a measurable advantage favoring the differentiated at every level except in the first quartile, at which level they exactly paralleled the ratios of the undifferentiated.

The educational quotients of the entire group presented no more striking difference than did the accomplishment ratios. The summary of all grades showed a mean educational quotient of 102.9 ± .3 for the undifferentiated and 102.0 ± .4 for the differentiated. Thus the reader may see that the difference between the two groups was .9 with a probable error of ± .5. Although this meager difference favored the undifferentiated, it was much too small to be statistically significant.

TABLE XII

THIRD-QUARTILE ACCOMPLISHMENT RATIOS

Grade	Undifferentiated	Differentiated
9	102.5	104.9*
8	107.5*	106.0
7	107.4*	106.9
6	106.5	113.3*
5	108.5	109.5*
4	107.8	112.6*
3	107.6*	105.3
2	107.1*	99.2

*Indicates measurable superiority at the point exceeding seventy-five percent of all ratios.

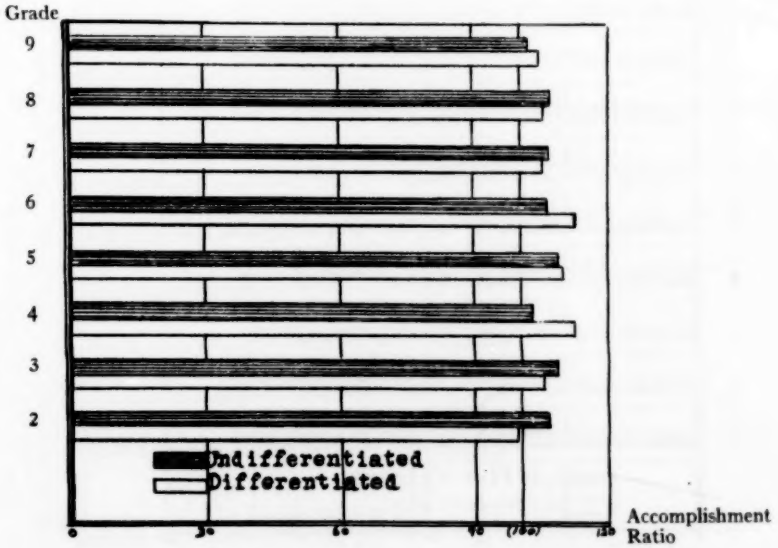


Figure 3.—Third-Quartile Accomplishment Ratios.

Since, as in the description of the accomplishment ratio, the various grades and levels must be studied to describe thoroughly the status, let us examine first the mean educational quotients, then the first quartile, and finally the third quartile.

The mean educational quotients for the various grades exhibited differences essentially like those of the accomplishment ratios. Grades two, three, seven, and eight showed differences favoring the undifferentiated groups, and grades four, five, six, and nine excelled in the differentiated groups. Only in one case was the difference as great as four times its probable error. This exception manifested itself in the second grade undifferentiated. In only two instances were the differences equal to more than three times the probable errors; namely, the third grade undifferentiated and the sixth grade differentiated. It follows, as a consequence, that in these four grades only were the differences large enough to be significant. The results are presented in Table XIII.

The mean educational quotients are shown in Figure 4, which presents an interesting contrast to Figure 1, on median accomplishment ratios. Generally the educational quotients were higher than the accomplishment ratios.

TABLE XIII

MEAN DIFFERENCES OF EDUCATIONAL QUOTIENTS

Grade	Undifferentiated	Differentiated
All	.9 ± .5	
9		1.3 ± 1.0
8	2.5 ± 1.3	
7	2.1 ± 1.1	
6		4.8 ± 1.4
5		2.0 ± 1.3
4		2.9 ± 1.4
3	5.6 ± 1.5	
2	9.5 ± 1.8	

The first-quartile educational quotients for the groups, as a whole, were higher for the undifferentiated pupils than for the differentiated. In five out of the eight grades studied the undifferentiated were superior at the first-quartile level. In grades four, six, and nine the differentiated excelled, but in the ninth grade only was the difference really significant. Table XIV shows the various first-quartile quotients for all the grades studied.

The first-quartile educational quotients were generally low. Only in the ninth-grade

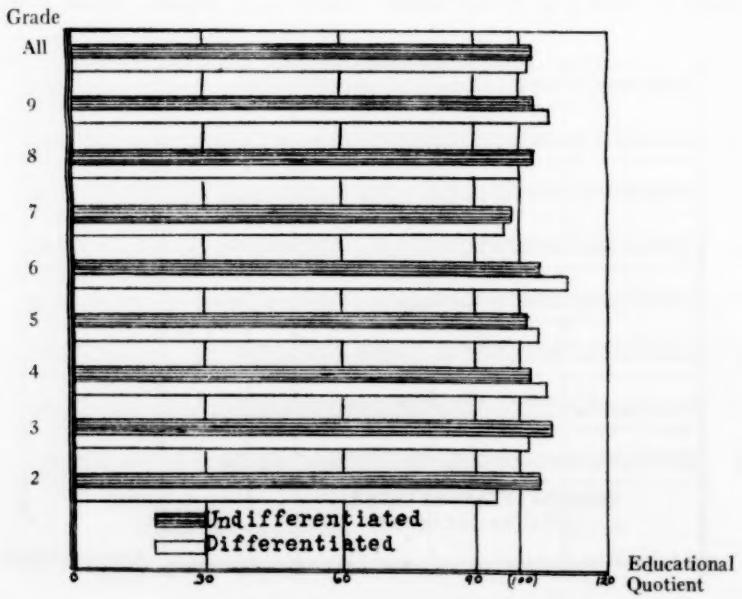


Figure 4.—Mean Educational Quotients.

TABLE XIV

FIRST-QUARTILE EDUCATIONAL QUOTIENTS

Grade	Undifferentiated	Differentiated
All	93.5*	91.5
9	94.3	102.5*
8	91.8*	88.9
7	89.9*	85.2
6	96.5	97.0*
5	93.1*	91.4
4	95.2	96.3*
3	99.6*	94.0
2	95.8*	83.5

*Indicates measurable superiority.

differentiated was the quotient equal to more than one hundred. The graph of Figure 5 shows the proximity of the various grades and groups to the one hundred line as well as the comparative achievement of the differentiated and the undifferentiated pupils.

The educational quotients of the third quartile, representing the pupils of high intelligence quotients, were highest in the undifferentiated groups for grades two, three, seven, eight, nine, and for the total. Among the differentiated groups, however, the highest value of all was found in the sixth grade. Grades four and five also excelled quite noticeably in the differentiated groups. Table XV presents these facts in concise tabular form, similar to the tables previously studied.

All the grades achieved educational quotients beyond one hundred; in most instances the values may be considered relatively high, in view of the fact that these same pupils had accomplishment ratios of less than one hundred, with very few exceptions. Figure 6 presents these third-quartile educational quotients graphically.

TABLE XV

THIRD-QUARTILE EDUCATIONAL QUOTIENTS

Grade	Undifferentiated	Differentiated
All	113.3*	112.1
9	115.2*	113.0
8	113.1*	111.3
7	109.8*	107.8
6	113.6	122.3*
5	111.5	115.7*
4	111.8	115.9*
3	117.3*	109.4
2	114.8*	105.8

*Indicates measurable superiority.

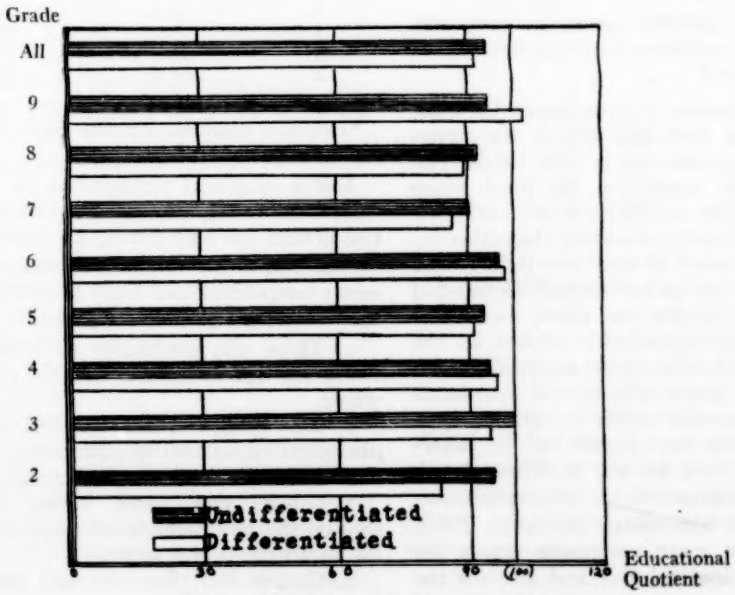


Figure 5.—First-Quartile Educational Quotients.

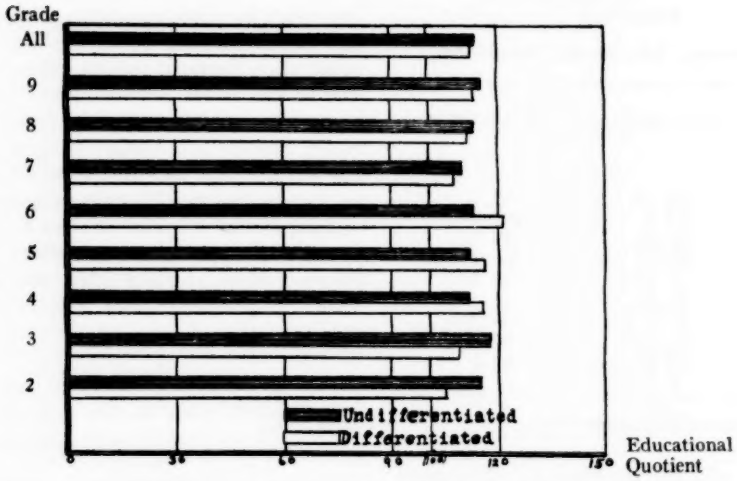


Figure 6.—Third-Quartile Educational Quotients.

In a summary analysis of the educational quotients, the following findings are worthy of notice:

1. The groups as a whole were approximately equal, the mean quotients being 102.9 ± .3 for the undifferentiated and 102.0 ± .4 for the differentiated.

2. The first-quartile quotients were higher for the undifferentiated than for the differentiated in five out of the eight analyzed.

3. The third quartile quotients were also higher for the undifferentiated in five out of the eight analyzed.

As a third measure of achievement the composite score was used, but only in the means for the various grades and in their totals. The mean composite scores for the total group were 76.2 for the undifferentiated and 75.2 for the differentiated, presenting the rather insignificant difference of only one point. The mean composite scores for the undifferentiated were superior in grades two, three, seven, and eight. A similar superiority existed in the identical grades for the mean accomplishment ratios and the mean educational quotients. The mean composite scores for grades four, five, six, and nine were higher for the differentiated pupils than for the undifferentiated. This fact also obtained in the accomplishment ratios and in the educational quotients. Table XVI shows the mean composite scores for grades two to nine inclusive and also for the total.

TABLE XVI

MEAN COMPOSITE SCORES

Grade	Undifferentiated	Differentiated
All	76.2*	75.2
9	98.7	102.3*
8	99.3*	90.5
7	82.6*	80.9
6	78.0	82.1*
5	65.5	71.4*
4	55.2	56.5*
3	44.8*	39.2
2	28.5*	21.4

*Indicates measurable superiority.

For a graphical presentation of the mean composite scores, showing the standing of the two groups for each grade, see Figure 7.

The additional evidence resulting from the mean-composite-score study is incorporated in the following significant statements:

1. There was practically no difference between the means representing the groups as a whole.

2. The grades which excelled in accomplishment ratios and in educational quotients also excelled in the mean-composite scores.

3. Grades two, three, seven, and eight showed the undifferentiated pupils superior in the mean composite scores.

4. Grades four, five, six, and nine showed the differentiated groups superior.

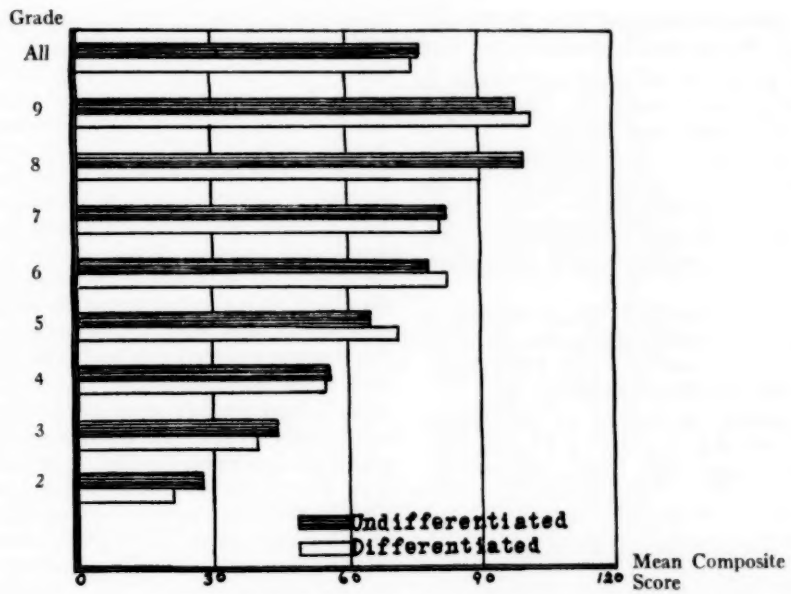


Figure 7.—Mean Composite Scores.

As a final step in the analysis of the two groups, the study turns now to a measure of relationship. The correlation between the intelligence quotients and the educational quotients presents a choice bit of evidence. For the entire group the coefficients were .73 ± .01 for the undifferentiated and .61 ± .015 for the differentiated, with a consequent difference of .12 ± .018 in favor of the former. Such variation is decidedly significant. Examining the individual grades, one finds the same relationship prevailing throughout. In every grade, except the second and the eighth, the coefficients were higher for the undifferentiated group. In the second grade, while the coefficient was higher for the differentiated, both coefficients were so low and the probable errors so high that the difference between the two was .10 ± .11, a difference entirely insignificant inasmuch as it is just about equal to its probable error. In the eighth grade both coefficients were exactly equal. For a tabular presentation see Table XVII. All of these correlations are graphically presented in the Appendix, with both regression lines superimposed upon the scatter diagrams for the individual grades and for the totals. The conclusion for these measures of relationship definitely in favor of the undifferentiated group may be summarized thus: The coefficients of correlation show that one can rather definitely predict the educational quotients from the intelligence quotients for the pupils in the undifferentiated group as a whole; further, one can predict them for most of the grades, with the exceptions of the second and the eighth grades, in which the prediction possibilities are essentially equal for both groups.

TABLE XVII

COEFFICIENTS OF CORRELATION BETWEEN I.Q.'S AND E.Q.'S

Grade	Undifferentiated	Differentiated
All	.73 ± .01*	.61 ± .015
9	.76 ± .02*	.71 ± .03
8	.79 ± .02	.79 ± .02
7	.76 ± .02*	.74 ± .02
6	.79 ± .03*	.54 ± .05
5	.79 ± .03*	.60 ± .05
4	.75 ± .04*	.61 ± .05
3	.56 ± .06*	.34 ± .07
2	.28 ± .08	.38 ± .07*

*Indicates measurable superiority.

The observations listed below conclude the findings of the four-measure test applied to

both the differentiated and the undifferentiated pupils in the grades studied.

1. Both groups were essentially equal in educational accomplishment, as measured by the mean accomplishment ratio.

2. The two groups were about equal in educational achievement, as measured by the mean educational quotient.

3. The groups differed to the extent of one point only, as measured by the mean composite score.

4. The undifferentiated group showed a higher degree of relationship between the intelligence quotient and the educational quotient than did the differentiated group.

5. The quartile deviations of the educational quotients were 9.9 and 10.3 for the undifferentiated and the differentiated respectively.

Interpretation of Results

In the description of the data thus far presented, it is evident that the main observations, although they recur sufficiently often to be called central tendencies, frequently do not present the entire picture. In generalizing upon the findings already presented, the writer has taken this fact into consideration and invites his readers to practice similar precautions. The interpretation of the results will follow the outline previously given: first, the accomplishment ratio; second, the educational quotient; third, the raw composite score; and fourth, the correlation between the intelligence quotients and the educational quotients.

Differentiation, in so far as this investigation is concerned, does not materially benefit pupils in their achievement of high accomplishment ratios. From all available evidence, the pupils of high intelligence do slightly better when in classes where differentiation is not practiced than when in differentiated groups. Those pupils who are only average in ability do as well when segregated as when in mixed classes. The dull pupils in differentiated classes, however, excel in comparison with their mates in undifferentiated classes. This superiority or inferiority of accomplishment ratios must be interpreted to mean that accomplishment is the quotient of achievement and of intelligence and not the total picture of achievement. Throughout the various grades, the mean accomplishment was very much the same for both groups; when, however, any superiority existed, it existed measurably in favor of the undifferentiated pupils. This fact seems to indicate, when considered generally, that for accomplishment the 787 pupils in the mixed classes had a slight measurable advantage over the segregated pupils. These results agree almost exactly with those found by Billett to the effect that the bright pupils in the differentiated groups were slightly disadvantaged, the average enjoyed small advantage, and the dull, a marked advantage.[43]

As a final interpretation of the accomplishment study, it is important to record that the pupils of the highest ability invariably made high scores but low accomplishment ratios, that is, their scores were not as high as one would expect them to be from their mental ability; hence their achievement in relation to their intelligence was low. This condition prevailed in both groups, although, as already indicated, the bright pupils had better chances of accomplishment in the undifferentiated classes.

The educational quotients of the pupils in the undifferentiated group in general were superior to those of the differentiated group. In the first quartile, or among those pupils having low intelligence quotients, the undifferentiated were decidedly superior. The pupils of average ability on the whole did as well in mixed classes as in differentiated classes. In those grades where the undifferentiated were superior, they were very decidedly superior, although the differentiated showed only a slight measurable advantage in one grade. To be sure, when all grades were taken together no group was really superior; but to say, therefore, that no difference existed at all would be incorrect. It would be more nearly correct to say that in most cases the undifferentiated pupils of average ability had the advantage. The undifferentiated third-quartile pupils of high ability excelled in five out of eight grades and in the total. Needless to add, that so far as the educational quotients were concerned, the undifferentiated pupils in this study had the advantage. This finding agrees with the results of the educational-accomplishment study except among pupils of

[43] R. O. Billett, "A Controlled Experiment to Determine the Advantages of Homogeneous Grouping," *Educational Research Bulletin*, VII (April 4–May 2, 1928).

low ability, who apparently had better chances of accomplishment when differentiated, but measurable disadvantage in the educational quotient.

The mean composite scores were approximately the same for the total in both groups. There was only one point of difference in favor of the undifferentiated group. In four grades the differentiated pupils were superior to a slight degree, while the undifferentiated were superior in four grades to a greater degree. Even though the mean difference between the groups was small, the undifferentiated must be regarded as slightly superior to the differentiated. This finding accords with the previous findings of the survey. Considering the composite score as the source for the achievement data used for the other measures, one correctly concludes that for group measurement the raw composite score will give approximately the same prediction of achievement as do the more complicated measures used. This fact, however, becomes evident only after all measures have been applied and thoroughly described. Knowing a pupil's score, it appears, will go a long way in the accurate description of his educational status, but to know his intelligence quotient will in a great many cases prove adequate for achievement description and accomplishment description.

Within the scope of this study there is a greater relationship between intelligence quotients and educational quotients for the undifferentiated pupils than for the differentiated. Having determined the intelligence quotient for a pupil in mixed classes, it is possible to predict his educational quotient to a greater degree of accuracy than if he were in differentiated classes. Whether one considers accomplishment, educational quotients, composite scores, or relationship between intelligence and educational quotients, makes little difference so far as this study is concerned. Virtually, the undifferentiated were slightly superior to the differentiated except for a few isolated grades.

Pertinent Applications

The several conclusions of this survey, arrived at through the tedious methods of research and statistical investigation, may help the anxious school administrator in whose mind sectioning (its bases and advantages) is still problematical. They should help to answer more definitely for school men generally whether differentiation should be practiced more widely than at the present time. The several results in the tables prepared could well be used as temporary norms, although the writer does not claim that such norms have been permanently established. The various scatter diagrams and regression-line equations could be used for prediction purposes. Each of these possible outcomes of the study will be considered briefly.

From the presentation of the facts herein analyzed and interpreted, it appears that differentiation within the scope of this study did not materially improve educational achievement. The writer does not imply, however, that under no conditions could differentiation affect educational achievement and improvement. According to the studies previously made and, again, in this investigation, the conclusion seems to hold that not differentiation alone but other curricular and instructional measures together with differentiation bring desired results in educational improvement. Thus far, therefore, the survey is but a verification of previous researches with very little additional evidence.

The first-quartile accomplishment ratios should prove helpful as temporary norms for bright pupils. An interesting phase of high intelligence manifested itself at this juncture of the research. Educators will do well to look with amazement and concern at the rather contradictory finding that bright pupils, who should, according to current educational opinion, make high accomplishment ratios, invariably make low accomplishment ratios. Administrators might find it worthwhile to experiment with a group of pupils having intelligence quotients of more than 120, with the purpose of raising their first-quartile accomplishment so as to make it comparable with the third quartile educational quotients. By what means of intellectual stimulation this could be accomplished, only such an experiment could reveal. The mean and third-quartile accomplishment ratios might serve as a check on the experiment suggested or they too might be used as temporary class norms for class evaluation. These values must not be considered as valid by themselves for individual diagnosis; if, on the other hand, they are used with a full appreciation of their shortcomings, they can prove exceedingly beneficial.

The educational quotients of the first quartile, the mean, and the third quartile could be used as temporary norms for dull, average, and bright pupils respectively. The composite scores are given only for the average of each of the eight grades; hence, they could be used only to rate the standing of a particular grade. They too are merely supplementary to the norms already established and, therefore, are most beneficial when used in conjunction with the other measures as herein described.

Finally the regression equations, as found in the Appendix, can be used to predict educational quotients from intelligence quotients. The regression lines, as superimposed on the scatter diagram, would serve for approximate quotient evaluations and could be used successfully from the fourth grade to the ninth. The reader must remember that the low correlation between the intelligence quotients and the educational quotients makes the relationship of the second and third grade quotients of doubtful value.

A thorough perusal of the data presented throughout the description of the various grades will suggest other applications to the interested administrator, but those here suggested seem to be the most obvious and also the most timely.

Conclusion

The conclusions of this investigation will include a summary of the study, a statement regarding the limitations of the study, and several suggestions for further investigation.

A. *The Problem of the Investigation*
   1. Do pupils grouped according to ability show greater progress in educational achievement than those taught in the traditional way, if the factors of intelligence and grade-placement are balanced for both groups?

B. *Procedure and Development*
   1. Eleven schools participated in this investigation, four practicing differentiation and seven undifferentiation.
   2. These schools provided 2,325 children from grades two to nine inclusive. In undifferentiated classes there were 1,163 pupils and 1,162 in differentiated classes.
   3. The number of pupils was reduced to 787 for each group by matching for grade and intelligence.
   4. The eleven schools furnished the intelligence quotients.
   5. The New Stanford Achievement Tests were used to measure the educational achievement.
   6. The test data and the intelligence quotients were the sources from which were determined the accomplishment ratio, the educational quotient, the composite score, and the correlation between the intelligence quotients and the educational quotients.
   7. The results were tabulated and described for each grade of each group and for the total.

C. *Summary of Grades and Total*
   1. Second Grade
      a. The undifferentiated group was superior in educational accomplishment for the first quartile, the mean, and for the third quartile.
      b. The undifferentiated group was superior at every level for the educational quotient.
      c. The mean composite score was higher for the undifferentiated group than for the differentiated.
      d. The low correlation existing between the intelligence quotients and the educational quotients for both groups makes it unsafe to predict educational quotients from intelligence quotients.
   2. Third Grade
      a. The undifferentiated group excelled in educational accomplishment at every level.
      b. The undifferentiated group had higher educational quotients at every level.
      c. The mean composite score favored the undifferentiated.
      d. The correlation between the intelligence quotients and the educational quotients was higher for the undiffer-

entiated but the difference between the coefficients was negligible.

3. Fourth Grade
   a. Both groups were about equal in educational accomplishment for the first quartile and the median, but in the third quartile the differentiated group was definitely superior.
   b. The two groups were about equal in educational quotients. All, however, favored the differentiated group.
   c. Both groups were about equal in the mean composite score.
   d. The correlation between the intelligence quotients was .75 and .61 for the undifferentiated and the differentiated respectively.
4. Fifth Grade
   a. Both groups were approximately equal in educational accomplishment and educational quotients, the greatest difference being in the third-quartile where educational quotients favored the differentiated group.
   b. The differentiated group was decidedly superior in the composite score.
   c. The coefficients of correlation were .79 and .60 for the undifferentiated and the differentiated respectively. The difference between the coefficients was .19 ± .05 in favor of the undifferentiated.
5. Sixth Grade
   a. The differentiated excelled at every level in educational accomplishment and in educational quotients.
   b. The differentiated group was superior in composite scores.
   c. The relationship between intelligence quotients and educational quotients was highest for the undifferentiated.
6. Seventh Grade
   a. Both groups were essentially equal in accomplishment and educational quotients.
   b. Both groups were about equal in mean composite scores.
   c. The correlation between intelligence and achievement was .76 for the undifferentiated and .74 for the differentiated.
7. Eighth Grade
   a. The undifferentiated group was superior in educational accomplishment, in educational quotients, and in mean composite scores.
   b. The coefficient of correlation between intelligence and achievement was .79 for both groups.
8. Ninth Grade
   a. The differentiated group was superior at every level in educational accomplishment.
   b. The differentiated slightly excelled in educational quotients for the first quartile and for the mean. The undifferentiated excelled in educational quotients for the third quartile.
   c. The differentiated excelled in composite scores.
   d. The relationship between intelligence and achievement was .76 and .71 for the undifferentiated and the differentiated respectively.
9. All Grades
   a. Both groups were about equal in mean composite scores.
   b. Both groups were almost equal in educational quotients. The undifferentiated were slightly superior at every level.
   c. The groups differed only one point in the mean composite score; this difference was in favor of the undifferentiated group.

d. The correlation between the intelligence quotients and the educational quotients was .73 for the undifferentiated and .61 for the differentiated. The difference between the coefficients was .12 ± .018 in favor of the undifferentiated group.

D. *Limitations of the Study*

1. The administering of the tests was left to the principals of the cooperating schools and no attempt was made to check their methods.
2. The teacher factor, as is always the case, was a variable, and, although practically all the teaching was supervised, the teacher ratings were not considered.
3. There were more schools contributing undifferentiated pupils than differentiated.
4. The number of pupils included in this study might well have been larger.
5. Curricular adaptations made because of groupings were not checked.

E. *Suggestions for Further Investigation*

1. Does there exist between the intelligence quotients and the educational quotients for pupils of grades 7, 8, and junior high school a higher relationship than in an 8–4 organization?
2. What system of instruction, thus far known, most noticeably challenges the pupils having intelligence quotients of 120 and over to work to capacity?
3. Can the educational philosophy of the principal of a school be measured objectively by testing the pupils under his administration?

F. *Agreement with Previous Researches in Psychology*

The findings of this study pertaining to accomplishment ratios agree with previous researches, by indicating that the higher the intelligence quotient is, the lower will be the accomplishment ratio.

Bibliography

1. Adler, Martha. "Mental Tests Used as a Basis for the Classification of School Children," *Journal of Educational Psychology*, V (January, 1914), 22–28.
2. Alexander, J. M. "Binet-Simon Tests in Practical Use in the Public Schools of Hinsdale, Illinois," *Elementary School Journal*, XXI (October, 1920), 146–48.
3. Armentrout, W. D. "Classification of Junior Highschool Pupils by the Otis Scale," *Journal of Educational Psychology*, XI (March, 1920), 165–68.
4. Arnold, H. J. "Seashore's Plan of Sectioning on the Basis of Ability as a Motivation Device," *Psychology Monthly*, XXXIX (1928), 102–9.
5. Arthur, Grace "An Application of Intelligence Tests to the Problem of School Retardation," *School and Society*, X (November 22, 1919), 614–20.
6. Barnard, F. J. "Let Pupils Be So Classified as to Allow Unrestricted Progress or Unlimited Time According to Ability," *Proceedings of the National Educational Association*, XXXVIII (1899).
8. Barthelmess, Harriet M., and Boyer, Philip A. "An Evaluation of Ability Grouping," *Journal of Educational Research*, XXVI (1932), 284–94.
9. Beeson, M. F., and Tope, Richard. "The Educational and Accomplishment Quotients as an Aid in the Classification of Pupils," *Journal of Educational Research*, IX (April, 1924), 281–92.
10. Berry, C. S. "Classification by Tests of Intelligence of Ten Thousand First-grade Pupils," *Journal of Educational Research*, VI (October, 1922), 185–230.
11. Billett, R. O. "A Controlled Experiment to Determine the Advantages of Homogeneous Grouping," *Educational Research Bulletin*, VII (April 4–May 2, 1928).
12. Bonar, H. S. "Ability Grouping in the First Grade," *Elementary School Journal*, XXIX (May, 1929), 703–6.
13. Branson, E. P. "An Experiment in Arranging Highschool Sections on the Basis of General Ability," *Journal of Educational Research*, III (January, 1921), 53–55.

14. Briggs, T. H. "Provisions for Abilities by Means of Homogeneous Grouping," *Third Yearbook of the National Association of Secondary School Principals.* Menasha, Wisconsin: George Banta Publishing Co., 1919. Pp. xxxi + 87.

15. Brink, Laurence B. "The Fallacy of Ability Grouping," *School and Society,* XXXV (March 26, 1932), 427–29.

16. Brooks, F. D. "Sectioning Junior High School Pupils by Tests and School Marks," *Journal of Educational Research,* XII (December, 1925), 359–69.

17. Buckingham, B. R. "The Validity of the Conclusions from the Statistical Data (of Sec. III, Twenty-fourth Yearbook)," *Twenty-fourth Yearbook of the National Society for the Study of Education,* pp. 216–21. Bloomington, Illinois: Public School Publishing Co., 1925.

18. Burtt, H. E., Chassell, L. M., and Hatch, E. M. "Efficiency of Instruction in Unselected Sections in Elementary Psychology Compared with That in Sections Selected on Basis of Intelligence Tests," *Journal of Educational Psychology,* XIV (March, 1923), 154–61.

19. Bush, Emma D. "An Experiment in Varying Methods of Teaching Bright and Dull Sections of the Seventh Grade," *Journal of Educational Research,* VII (1923), 113–21.

20. Callihan, T. W. "An Experiment in the Use of Intelligence Tests as a Basis for Proper Grouping and Promotions in the Eighth Grade," *Elementary School Journal,* XXI (February, 1921), 465–69.

21. Clerk, F. E. "The Arlington Plan of Grouping Pupils According to Ability in the Arlington High School, Arlington, Mass.," *School Review,* XXV (January, 1917), 26–47.

22. Constance, Jennie M., and Hanna, Joseph. "An Experiment in Sectioning Freshmen English," *Educational Review,* LXVII (March, 1924), 150–52.

23. Cook, R. R. "A Study of the Results of Homogeneous Grouping of Abilities in High School Classes," *Twenty-third Yearbook of the National Society for the Study of Education.* Bloomington, Illinois: Public School Publishing Co., 1924.

24. Courtis, Stuart A. "Ability Grouping in Detroit Schools," *Twenty-fourth Yearbook of the National Society for the Study of Education,* pp. 44–47. Bloomington, Illinois: Public School Publishing Co., 1925.

25. Dawson, C. D. "Classification of Kindergarten Children for First Grade by Means of the Binet Scale," *Journal of Educational Research,* VI (December, 1922), 412–22.

26. Douglass, Harl R. "Certain Aspects of the Problem of Where We Stand With Reference to the Practicability of Grouping," *Journal of Educational Research,* XXVI (1933), 344–53.

27. Dvorak, August, and Rae, J. J. "A Comparison of the Achievements of Superior Children in Segregated and Unsegregated First-grade classes," *Elementary School Journal,* XXIX (January, 1929), 380–86.

28. Engel, Anna M. "Comparison of Class Ratings of Pupils in Special Advanced Classes with Accelerated Pupils in Regular Classes in the Detroit Public Schools," *Twenty-third Yearbook of the National Society for the Study of Education, Part I,* pp. 297–301. Bloomington, Illinois: Public School Publishing Co., 1924.

29. Feingold, G. A. "The Sectioning of High-school Classes on the Basis of Intelligence," *Educational Administration and Supervision,* IX (October, November, 1923), 399–415, 467–86.

30. Feingold, G. A. "Views of Teachers on Sectioning of High School Classes on the Basis of Intelligence," *Educational Administration and Supervision,* IX (November, 1923), 467–86.

31. Fleming, C. W. *A Detailed Analysis of Achievement in the High School.* Teachers College Contributions to Education, No. 196. New York: Teachers College, Columbia University, 1925. xi + 209.

32. Freeman, F. N. "Introduction and Summary to Chapter XII, on Miscellaneous Experimental and Statistical Studies of Gifted Children," *Twenty-third Yearbook of the National Society*

for the Study of Education. Bloomington, Illinois: Public School Publishing Co., 1924.

33. Freeman, F. N. "Sorting the Students," *Educational Review*, LXVIII (November, 1924), 169–74.

34. Glass, J. M. "Classification of Pupils in Ability Groups," *School Review*, XXVIII (September, 1920), 496–508.

35. Greenwell, W. M. "The Classification of Children in the Lincoln School, Oakland, California," *Bulletin of the Department of Elementary School Principals, Second Yearbook*, July, 1923.

36. Hughes, J. N. "A Study of Intelligence and of the Training of Teachers as Factors Conditioning the Achievement of Pupils, II," *School Review*, XXXIII (April, 1925), 292–302.

37. Hughes, W. Hardin. "Organized Personnel Research and Its Bearing on High-School Problems," *Journal of Educational Research*, X (December, 1924), 386–98.

38. Kefauver, Grayson N. "The Validity of Bases for Forming Ability Groups," *Teachers College Record*, XXI (November, 1929), 99–115.

39. Kelley, T. L. *Educational Guidance*. Teachers College Contributions to Education, No. 71. New York: Teachers College, Columbia University, 1919.

40. McCall, Wm. A. "A Comparison of the Educational Progress of Bright Pupils in Accelerated and in Regular Classes," *Twenty-seventh Yearbook of the National Society for the Study of Education, Part II*. Bloomington, Illinois: Public School Publishing Co., 1928.

41. Maher, E. A. "An Experiment in Homogeneous Grouping," *Journal of Educational Method*, V (September, 1925), 25–28.

42. Miller, W. S. "General Intelligence Tests," *School Review*, XXVIII (February, 1920), 94.

43. Miller, W. S. "The Administrative Uses of Intelligence Tests in the High School," *Twenty-first Yearbook of the National Society for the Study of Education*, Part II. Bloomington, Illinois: Public School Publishing Co., 1922.

44. Moyer, Edward L. "A Study of the Effects of Classification by Intelligence Tests," *Twenty-third Yearbook of the National Society for the Study of Education*. Bloomington, Illinois: Public School Publishing Co., 1924.

45. Otis, Arthur S. "Ability Grouping," *School and Society*, XXXVI (July 23, 1932), 116–18.

46. Patrick, Mary L. "Some Attainments of Gifted Children in Segregated Classes at Louisville," *Twenty-third Yearbook of the National Society for the Study of Education*. Bloomington, Illinois: Public School Publishing Co., 1924.

47. Pierson, C. D., and Nettels, C. H. "A Study of High School Seniors to Determine Who Shall be Recommended to College," *School and Society*, XXVIII (August, 1928), 215–16.

48. Pintner, R., and Noble, H. "The Classification of School Children According to Mental Age," *Journal of Educational Research*, II (November, 1920), 713–28.

49. Poffenberger, A. L., and Carpenter, F. L. "Character Traits in School Success," *Journal of Experimental Psychology*, VII (February, 1924), 67–74.

50. Pressey, S. L. "An Attempt to Measure the Comparative Importance of General Intelligence and General Character Traits in Contributing to Success in School," *Elementary School Journal*, (November, 1920), 220–27.

51. Price, E. D. "A Plan of Classifying Pupils," *Journal of Educational Research*, XII (December, 1925), 341–48.

52. Purdom, T. L. *The Value of Homogeneous Grouping*. Baltimore: Warwick and York, 1929. Pp. 96.

53. Rainey, H. P., and Anderson, H. "An Experiment in Classifying High School Pupils on the Basis of Achievement," *Educational Administration and Supervision*, XIII (October, 1927), 528–44.

54. Reeve, W. D. "Homogeneous Grouping of High School Students by Means of Psychological Tests," *Fourth Yearbook of the National Association of Secondary School Principals*, 1920.

55. Rowland, S. V. "Changes in Educational Procedure Growing Out of the Group-

ing of Pupils by Ability," *Tenth Annual School Men's Week Proceedings*, pp. 64–69. University of Pennsylvania, Bulletin No. 23, XXXVIII, June 9, 1923.

56. Ryan, H. H., and Crecelius, P. *Ability Grouping in the Junior High School.* New York: Harcourt Brace, 1927. Pp. 223.

57. Seashore, C. E. "Sectioning Classes on the Basis of Ability," *School and Society*, XV (1922), 353–58.

58. Senour, A. C. "Some Influences of the Multiple Track Organization," *Bulletin of the School of Education, Indiana University*, III (May, 1927).

59. Shields, J. M. "Teaching Reading Through Ability Grouping," *Journal of Educational Method*, VII (September–October, 1927), 7–10.

60. Smead, A. E. "Experiment in Classification of Ninth Grade Pupils According to Ability," *School Review*, XXV (November, 1927), 673–75.

61. Stedman, Lulu M. *The Education of Gifted Children.* Yonkers-on-Hudson: World Book Co., 1924. Pp. viii + 190.

62. Stone, C. L. "Disparity Between Intelligence and Scholarship," *Journal of Educational Psychology*, XIII (April, 1922), 241–44.

63. Sullivan, Elizabeth T. "Data on Ability Grouping from Los Angeles," *Twenty-fourth Yearbook of the National Society for the Study of Education*, pp. 148–51. Bloomington, Illinois: Public School Publishing Co., 1925.

64. Symonds, P. M. *Measurement in Secondary Education.* New York: The MacMillan Company, 1928.

65. Terman, L. M., and DeVoss, James C. "The Educational Achievements of Gifted Children," *Twenty-Third Yearbook of the National Society for the Study of Education*, Chapter X. Bloomington, Illinois: Public School Publishing Co., 1924.

66. Theisen, W. W. "The Relative Progress of VIIB Groups Sectioned on the Basis of Ability," *Journal of Educational Research*, V (April, 1922), 295–305.

67. Torgenson, T. L. "Is Classification by Mental Tests and Intelligence Quotients Worth While?" *Journal of Educational Research*, XIII (March, 1926), 171–80.

68. Turney, A. H. "The Status of Ability Grouping," *Educational Administration and Supervision*, XVII (January, February, 1931), 21–42, 110–27.

69. Van Wagenen, M. J. "Effect of Homogeneous Grouping upon the Quality of Work of Superior Children," *Journal of Educational Method*, VI (February, 1927), 240–47.

70. Ward, Mary A., Carter, Grace E., Holmes, Hilda M., and Anderson, Cecilia. "Data on Ability Grouping from the San Francisco State Teachers College," *Twenty-fourth Yearbook of the National Society for the Study of Education*, pp. 154–59. Bloomington, Illinois: Public School Publishing Co., 1925.

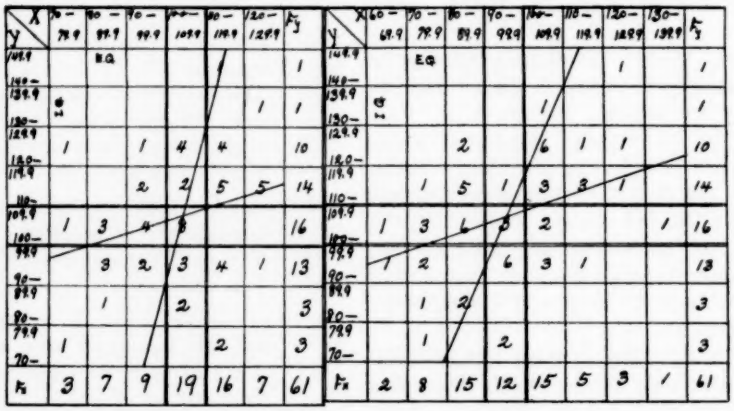
71. Worlton, J. T. "The Effect of Homogeneous Classification on Scholastic Achievement of Bright Pupils," *Elementary School Journal*, XXVIII (January, 1928), 336–45.

72. Worlton, J. T. "Why of Homogeneous Classification," *Elementary School Journal*, XXVII (December, 1926), 265–74.

Appendix

Summary of Correlation Charts and Regression Equations

Undifferentiated *Differentiated*



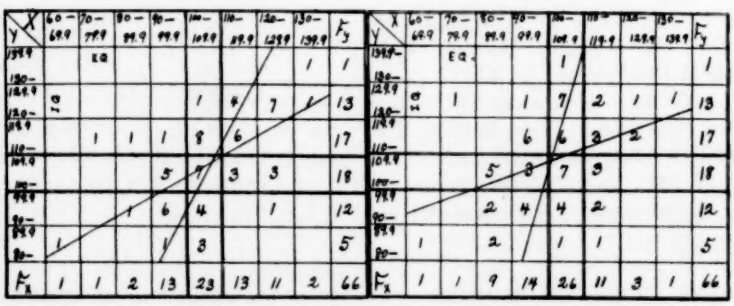
Y \ X	70 – 79.9	80 – 89.9	90 – 99.9	100 – 109.9	110 – 119.9	120 – 129.9	130 – 139.9	f_y
140 – 149.9		E.Q.			1		1	1
130 – 139.9	E.Q.					1	1	1
120 – 129.9	1		1	4	4		10	10
110 – 119.9			2	2	5	5	14	14
100 – 109.9	1	3	4	8			16	16
90 – 99.9		3	2	3	4	1	13	13
80 – 89.9		1		2			3	3
70 – 79.9	1				2		3	3
f_x	3	7	9	19	16	7	61	61

Y \ X	60 – 69.9	70 – 79.9	80 – 89.9	90 – 99.9	100 – 109.9	110 – 119.9	120 – 129.9	130 – 139.9	f_y
140 – 149.9		E.Q.				1			1
130 – 139.9	E.Q.				1				1
120 – 129.9			2		6	1	1		10
110 – 119.9		1	5	1	3	3	1		14
100 – 109.9	1	3	4	5	2			1	16
90 – 99.9	1	2		6	3	1			13
80 – 89.9		1	2						3
70 – 79.9		1		2					3
f_x	2	8	15	12	15	5	3	1	61

$r = .28 \pm .08$ $r = .38 \pm .07$

Figure 8.—Distributions Of Correlated Intelligence Quotients (Y) And Educational Quotients (X) For The Second Grade.

Undifferentiated *Differentiated*



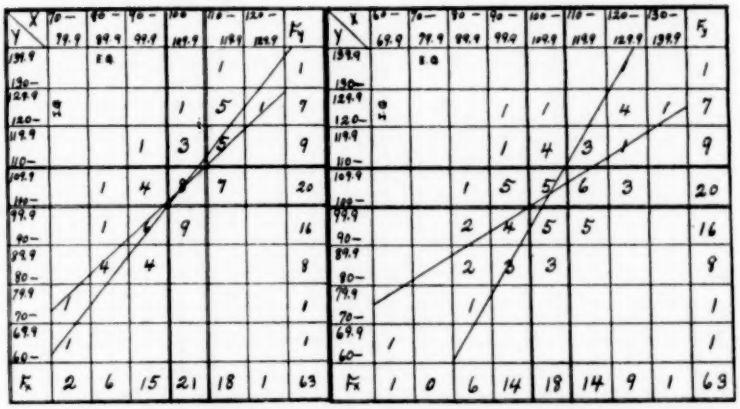
Y \ X	60 – 69.9	70 – 79.9	80 – 89.9	90 – 99.9	100 – 109.9	110 – 119.9	120 – 129.9	130 – 139.9	f_y
130 – 139.9		E.Q.						1	1
120 – 129.9	E.Q.				1	4	7	1	13
110 – 119.9		1	1	1	8	6			17
100 – 109.9				5	7	3	3		18
90 – 99.9			1	6	4		1		12
80 – 89.9	1			1	3				5
f_x	1	1	2	13	23	13	11	2	66

Y \ X	60 – 69.9	70 – 79.9	80 – 89.9	90 – 99.9	100 – 109.9	110 – 119.9	120 – 129.9	130 – 139.9	f_y
130 – 139.9		E.Q.			1				1
120 – 129.9	E.Q.	1		1	7	2	1	1	13
110 – 119.9				6	6	3	2		17
100 – 109.9			5	8	7	3			18
90 – 99.9			2	4	4	2			12
80 – 89.9	1		2		1	1			5
f_x	1	1	9	19	26	11	3	1	66

$r = .56 \pm .06$ $r = .34 \pm .07$

Figure 9.—Distributions Of Correlated Intelligence Quotients (Y) And Educational Quotients (X) For The Third Grade.

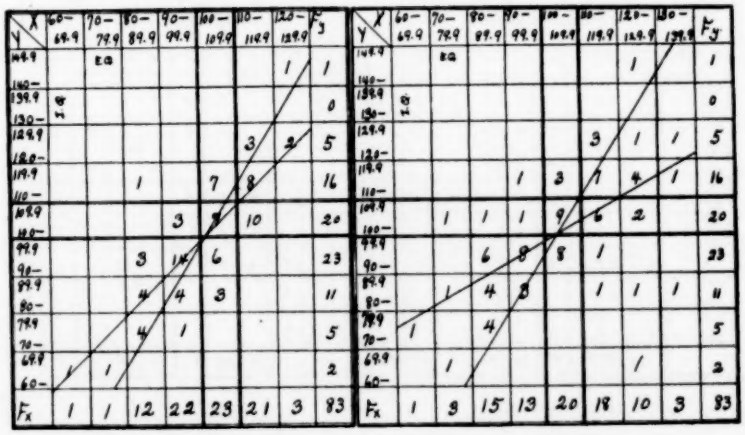
Undifferentiated *Differentiated*



r = .75 ± .04 r = .61 ± .05

Figure 10.—Distributions Of Correlated Intelligence Quotients (Y) And Educational Quotients (X) For The Fourth Grade.

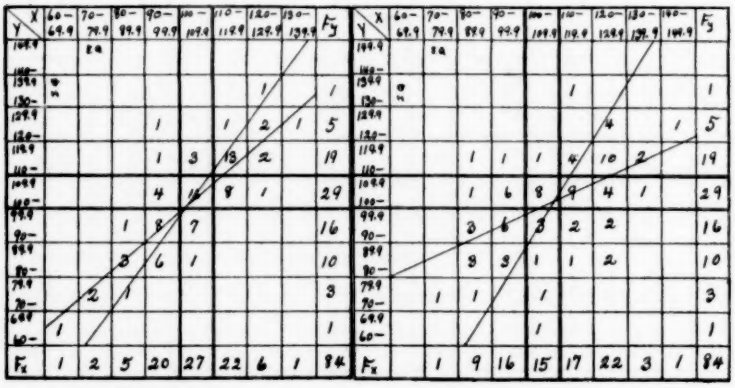
Undifferentiated *Differentiated*



r = .79 ± .03 r = .60 ± .05

Figure 11.—Distributions Of Correlated Intelligence Quotients (Y) And Educational Quotients (X) For The Fifth Grade.

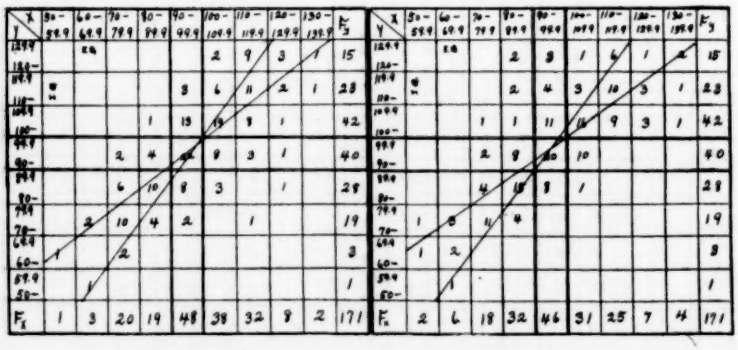
Undifferentiated *Differentiated*



r = .79 ± .03 r = .54 ± .05

Figure 12.—Distributions Of Correlated Intelligence Quotients (Y) And Educational Quotients (X) For The Sixth Grade.

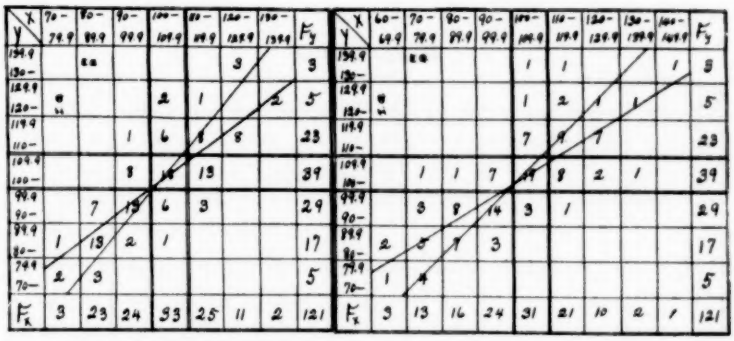
Undifferentiated *Differentiated*



r = .76 ± .02 r = .74 ± .02

Figure 13.—Distributions Of Correlated Intelligence Quotients (Y) And Educational Quotients (X) For The Seventh Grade.

Undifferentiated *Differentiated*



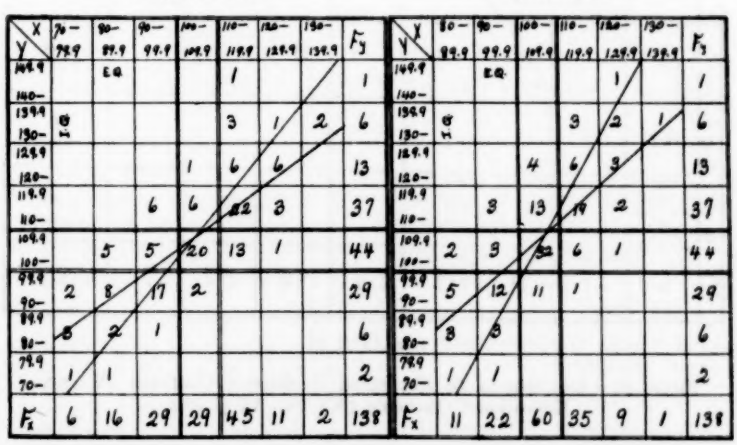
Y \ X	70–79.9	80–89.9	90–99.9	100–109.9	110–119.9	120–129.9	130–139.9	F_y
139.9 130–		EQ				3		3
129.9 120–	M			2	1		2	5
119.9 110–			1	6	8	8		23
109.9 100–			8	11	13			39
99.9 90–		7	13	6	3			29
89.9 80–	1	13	2	1				17
79.9 70–	2	3						5
F_x	3	23	24	33	25	11	2	121

Y \ X	60–69.9	70–79.9	80–89.9	90–99.9	100–109.9	110–119.9	120–129.9	130–139.9	140–149.9	F_y
139.9 130–		EQ			1	1			1	3
129.9 120–	M				1	2	1	1		5
119.9 110–					7	8	7			23
109.9 100–		1	1	7	11	8	2	1		39
99.9 90–		3	8	14	3	1				29
89.9 80–	2	3	7	3						17
79.9 70–	1	1								5
F_x	3	13	16	24	31	21	10	2	1	121

$r = .79 \pm .02$ $r = .79 \pm .02$

Figure 14.—Distributions Of Correlated Intelligence Quotients (Y) And Educational Quotients (X) For The Eighth Grade.

Undifferentiated *Differentiated*



Y \ X	70–79.9	80–89.9	90–99.9	100–109.9	110–119.9	120–129.9	130–139.9	F_y
149.9 140–		EQ			1			1
139.9 130–	M				3	1	2	6
129.9 120–				1	6	6		13
119.9 110–			6	6	22	3		37
109.9 100–		5	5	20	13	1		44
99.9 90–	2	8	17	2				29
89.9 80–	3	2	1					6
79.9 70–	1	1						2
F_x	6	16	29	29	45	11	2	138

Y \ X	80–89.9	90–99.9	100–109.9	110–119.9	120–129.9	130–139.9	F_y
149.9 140–		EQ			1		1
139.9 130–	M			3	2	1	6
129.9 120–			4	6	3		13
119.9 110–		3	13	19	2		37
109.9 100–	2	3	32	6	1		44
99.9 90–	5	12	11	1			29
89.9 80–	3	3					6
79.9 70–	1	1					2
F_x	11	22	60	35	9	1	138

$r = .76 \pm .02$ $r = .71 \pm .03$

Figure 15.—Distributions Of Correlated Intelligence Quotients (Y) And Educational Quotients (X) For The Ninth Grade.

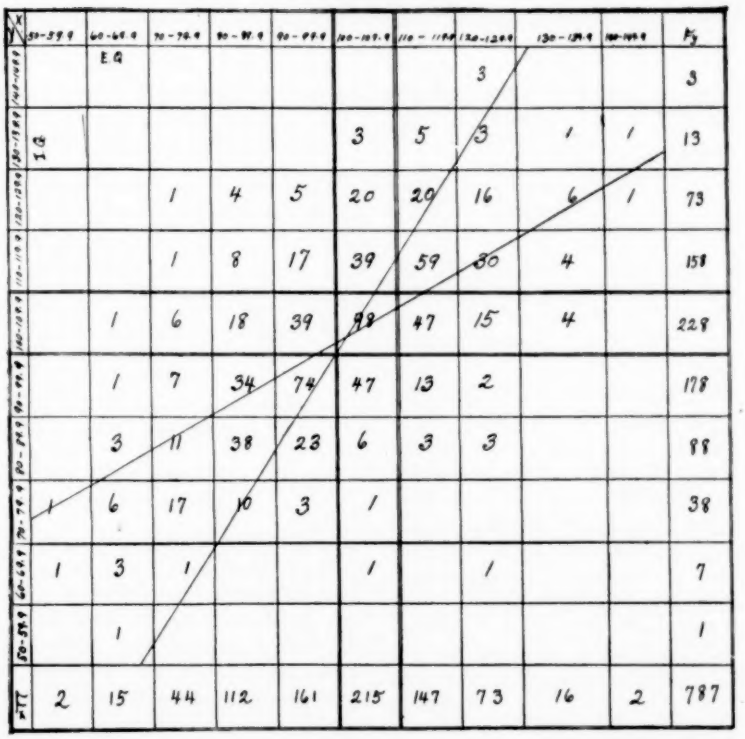
Undifferentiated

IQ \ EQ	60–69.9	70–79.9	80–89.9	90–99.9	100–109.9	110–119.9	120–129.9	130–139.9	Fy
140–149.9						2	1		3
130–139.9						4	6	3	13
120–129.9		1		2	11	33	21	5	73
110–119.9		1	2	15	41	78	20	1	158
100–109.9		1	10	46	103	62	6		228
90–99.9		4	28	88	45	10	3		178
80–89.9	1	10	37	25	14		1		88
70–79.9	2	16	14	3		3			38
60–69.9	2	4							7
50–59.9	1								1
fx	6	37	91	179	214	192	58	9	787

$r = .73 \pm .01$

Figure 16.—Distribution Of Correlated Intelligence Quotients (Y) And Educational Quotients (X) For All Grades.

Differentiated



Y \ X	50-59.9	60-69.9	70-79.9	80-89.9	90-99.9	100-109.9	110-119.9	120-129.9	130-139.9	140-149.9	f_y
140-149.9		E.G						3			3
130-139.9	I.G					3	5	3	1	1	13
120-129.9			1	4	5	20	20	16	6	1	73
110-119.9			1	8	17	39	59	30	4		158
100-109.9		1	6	18	39	98	47	15	4		228
90-99.9		1	7	34	74	47	13	2			178
80-89.9		3	11	38	23	6	3	3			88
70-79.9	1	6	17	10	3	1					38
60-69.9	1	3	1			1		1			7
50-59.9		1									1
f_x	2	15	44	112	161	215	147	73	16	2	787

$r = .61 \pm .015$

Figure 17.—Distribution Of Correlated Intelligence Quotients (Y) And Educational Quotients (X) For All Grades.

TABLE XVIII

UNDIFFERENTIATED GROUP EQUATIONS FOR REGRESSION LINES

Grade	Y on X	X on Y
2	$Y = .30X + 75.69$	$X = .26Y + 76.90$
3	$Y = .57X + 47.27$	$X = .55Y + 47.87$
4	$Y = .92X + 8.23$	$X = .81Y + 19.55$
5	$Y = .97X + 1.60$	$X = .64Y + 37.64$
6	$Y = .84X + 15.14$	$X = .75Y + 27.35$
7	$Y = .76X + 24.26$	$X = .76Y + 23.97$
8	$Y = .74X + 25.85$	$X = .84Y + 17.22$
9	$Y = .70X + 34.58$	$X = .83Y + 55.13$

For an approximate prediction of educational quotients from intelligence quotients, the scatter diagrams with the superimposed regression lines should prove helpful. For a more accurate prediction the regression line equations would serve better. In either case, however, it should be remembered that the coefficients were too low for much prediction, before the fourth grade.

TABLE XIX

DIFFERENTIATED GROUP EQUATIONS FOR REGRESSION LINES

Grade	Y on X	X on Y
2	$Y = .35X + 73.78$	$X = .42Y + 50.43$
3	$Y = .37X + 70.90$	$X = .31Y + 68.33$
4	$Y = .62X + 37.30$	$X = .60Y + 44.06$
5	$Y = .58X + 40.14$	$X = .62Y + 41.65$
6	$Y = .46X + 52.68$	$X = .63Y + 44.51$
7	$Y = .71X + 29.76$	$X = .77Y + 20.79$
8	$Y = .62X + 39.75$	$X = 1.00Y - 1.60$
9	$Y = .85X + 17.80$	$X = .58Y + 43.38$